GLOBAL PARTNERSHIP FOR YOUTH EMPLOYMENT

Mesurer le succès d'interventions axées sur les moyens de subsistance des jeunes

Un guide pratique du suivi et de l'évaluation

Kevin Hempel
Nathan Fiala

THE WORLD BANK

UCW Understanding Children's Work

international youth foundation

La version intégrale du guide, les notes individuelles et les ressources complémentaires sont disponibles en ligne à l'adresse suivante :
http://www.gpye.org/

©2011 Banque internationale pour la reconstruction et le développement/Banque mondiale
1818 H Street NW
Washington DC 20433
Téléphone : 202-473-1000
Internet : www.worldbank.org

Tous droits réservés.

Le présent ouvrage est le fruit du travail du personnel de la Banque internationale pour la reconstruction et le développement/la Banque mondiale. Les conclusions, interprétations et conclusions exprimées dans le présent document ne reflètent pas nécessairement les vues des Directeurs exécutifs de la Banque mondiale ou des gouvernements qu'ils représentent. La Banque mondiale ne garantit pas l'exactitude des données incluses dans ce document. Les frontières, couleurs, dénominations et autres informations sur n'importe quelle carte de cet ouvrage n'impliquent aucun jugement de la part de la Banque mondiale sur le statut légal d'un territoire ni la reconnaissance ou l'acceptation de ces frontières.

Droits et permissions
Ce document peut être adapté ou reproduit gratuitement à des fins uniquement non-commerciales, à condition que la source soit clairement mentionnée. Veuillez envoyer une copie de la partie du document reproduit à Kevin Hempel (khempel@worldbank.org).

Citation suggérée :
Hempel, Kevin, et Nathan Fiala, *Mesurer la réussite des interventions auprès des jeunes : Guide pratique sur le suivi et l'évaluation.* Washington, D.C.: Partenariat mondial pour l'emploi des jeunes, 2011, http://www.gpye.org/measuring-success-youth-livelihood-interventions

Couverture : Kathryn Werthman, Fondation internationale pour la jeunesse (IYF)
Photo : Nathan Fiala
Photo de la page 26 : Courtoisie de l'Organisation internationale du travail (OIT) / Le Réseau pour l'emploi des jeunes (YEN),
toutes les autres photos à l'intérieur du document sont la courtoisie de la Fondation internationale pour la jeunesse (IYF).

Nous vous saurions gré de nous faire part de vos commentaires pour améliorer le présent guide.
Veuillez adresser vos observations et suggestions à :

Kevin Hempel
khempel@worldbank.org

Nathan Fiala
nfiala@diw.de

Mesurer le succès d'interventions axées sur les moyens de subsistance des jeunes

Un guide pratique du suivi et de l'évaluation

Kevin Hempel
Nathan Fiala

L'apprentissage n'est jamais le fruit du hasard.
Il faut le rechercher avec ardeur et le suivre avec diligence.

— Abigail Adams

Table des matières

AVANT-PROPOS xi

REMERCIEMENTS xiii

INTRODUCTION xv
- Objet xv
- Destinataires xv
- Objectif xv
- But du guide xvi
- Études de cas xvii
- Étude de cas du NUSAF : Historique xvii
- Brève description du guide et mode d'emploi xix
- Guide du lecteur xix

NOTE 1 : Pourquoi évaluer ? 3
- Gestion de projets 4
- Création de connaissances 5
- Obligation de rendre des comptes 6
- Crédibilité et durabilité 7
- Les résultats négatifs d'une évaluation ne sont pas nécessairement mauvais 8
- Points essentiels 8
- Étude de cas du NUSAF : pourquoi évaluer ? 9
- À lire 9

NOTE 2 : Examen de la conception d'un projet 11
- Analyse du problème : comprenons-nous le groupe cible et le contexte local ? 12
- Diagnostic : quels sont les éléments déterminants dans l'obtention de résultats pour les jeunes ? 14
- Objectifs et élaboration : Quel est notre but ? Comment l'atteindre ? 19
- Points essentiels 23
- Étude de cas du NUSAF : Examen de la conception du projet 23
- À lire 24

NOTE 3 : Mettre sur pied un système de suivi 27
- Pourquoi avons-nous besoin d'un système de suivi ? 28
- Définir la logique de l'intervention 28
- Définir les indicateurs clés, les instruments de collecte des données et les hypothèses 33
- Mise en place d'un système de suivi et d'établissement de rapports 44
- Points essentiels 47
- Étude de cas du NUSAF : système de suivi 48
- À lire 48

NOTE 4 : Choisir le bon type d'évaluation 51
- Quel est le but de l'évaluation ? 52
- Relier les questions d'évaluation à la procédure d'évaluation 54

Le contexte dans lequel se déroule notre programme correspond-il au type d'évaluation souhaité ? 57
Types de programmes qui justifient habituellement la conduite d'une évaluation d'impact 62
Points essentiels 65
Étude de cas du NUSAF : Décider de conduire ou non une évaluation d'impact 66
À lire 67

NOTE 5 : Prouver l'impact d'un programme 71
La difficulté de déterminer la cause des résultats observés 72
Qu'est-ce que l'« impact » exactement ? 72
Comment pouvons-nous estimer le contrefactuel ?? 73
Techniques pour constituer des groupes témoins de qualité 74
Contrefactuels contrefaits 76
Points essentiels 79
Étude de cas du NUSAF : identifier un contrefactuel 80
À lire 81

NOTE 6 : Choisir une méthode d'évaluation d'impact appropriée 83
Opérer un choix parmi les méthodes d'évaluation d'impact 84
1re méthode : la procédure par tirage au sort 86
2e méthode : procédure randomisée par étapes 91
3e méthode : promotion aléatoire 93
4e méthode : méthode par discontinuité de la régression 97
5e méthode : double différence 99
6e méthode : appariement 102
Combinaison de méthodes 105
Points essentiels 109
Étude de cas du NUSAF : le choix d'une procédure par tirage au sort 110
À lire 110

NOTE 7 : Guide par étape d'une évaluation d'impact 113
Préparer l'évaluation d'impact 114
Définir le calendrier et le budget 115
Constituer une équipe d'évaluation 117
Élaborer un plan d'évaluation 119
Élaborer et conduire un outil d'enquête 124
Conduite d'une enquête de base et analyse 130
Conduite de l'enquête de suivi et analyse 133
La diffusion des résultats 137
Résolution des problèmes 139
Points essentiels 142
Étude de cas du NUSAF : Mise en œuvre de l'évaluation d'impact 143
À lire 143

NOTE 8 : Augmenter la pertinence de l'évaluation d'impact 145
Mesure de divers impacts 146
Utilisation de méthodes mixtes d'approches 148
Analyse coûts-avantages et coût-efficacité 150

Points essentiels 154
Étude de cas du NUSAF : Améliorer la pertinence d'une évaluation d'impact 155
À lire 156

Ouvrages cités 159

Bibliographie 167
Documentation sur le développement et l'emploi des jeunes 167
Documentation sur le suivi et l'évaluation 168
Bases de données de recherche 169
Journaux académiques 170
Bases de données existantes et permanentes sur les évaluations d'impact 171
Blogs 171
Renforcement des capacités 172
Sites Internet 172

ANNEXE 1. Exemples d'indicateurs pour les évaluations de la jeunesse 173

ANNEXE 2. Solutions pour diminuer les coûts ? 177

ANNEXE 3. Tests de vérification et de falsification 181

ENCADRÉS

Encadré 1.1 Données existantes sur l'emploi des jeunes 5
Encadré 1.2 Avantages d'une évaluation d'impact 8
Encadré 2.1 Évaluations sur un échantillon de jeunes et études de marché 14
Encadré 2.2 Le cadre MILES 15
Encadré 2.3 Définir un objectif de développement du projet 20
Encadré 4.1 Exemples d'évaluations par type 57
Encadré 4.2 Phases d'exécution d'un programme et stratégies d'évaluation appropriées 59
Encadré 4.3 Le manque de preuves quant aux programmes axés sur les moyens de subsistance des jeunes 63
Encadré 5.1 L'aléation est-elle éthique ? 75
Encadré 5.2 Exemples de cas où l'aléation n'est pas possible 76
Encadré 5.3 Validités interne et externe 76
Encadré 5.4 Exemples choisis d'évaluations non expérimentales 79
Encadré 6.1 Niveaux d'aléation 88
Encadré 6.2 Exemple de procédure par tirage au sort 90
Encadré 6.3 Exemple de procédure randomisée par étapes 93
Encadré 6.4 Conditions nécessaires pour que la promotion produise des estimations de l'impact valables 95
Encadré 6.5 Exemple de promotion aléatoire 96
Encadré 6.6 Exemple d'approche par discontinuité de la régression 99
Encadré 6.7 Exemple de méthode de la double différence 102
Encadré 6.8 Étapes à suivre pour appliquer une technique d'appariement 104
Encadré 6.9 Exemple d'appariement 105
Encadré 7.1 Aperçu d'un plan d'évaluation d'impact 119
Encadré 7.2 Les sources potentielles de données 122
Encadré 7.3 Facteurs influant sur la fiabilité des données au cours d'enquêtes s'adressant aux jeunes 124
Encadré 7.4 Aperçu du contenu d'un manuel d'enquête 126
Encadré 7.5 Formulaire-type d'un comité d'examen (IRB) 127
Encadré 7.6 Conseils sur la procédure d'approbation du IRB 128
Encadré 7.7 Exemple pratique : la protection des sujets humains 130
Encadré 7.8 Aperçu d'un rapport de base 132
Encadré 7.9 Questions courantes destinées à être ajoutées dans une enquête de suivi 134
Encadré 7.10 Exemple d'ajouts au rapport initial après l'enquête finale 135
Encadré 7.11 Le data mining (DM) 136
Encadré 7.12 Points de diffusion sélectionnés 138
Encadré 8.1 Exemple d'évaluation à partir de méthodes mixes 150

FIGURES

Figure 1.1 Avantages de l'évaluation **3**
Figure 1.2 Le cycle du projet **4**
Figure 1.3 Des données aux politiques : programmes de transferts conditionnels en espèces **6**
Figure 2.1 Environnements des jeunes **12**
Figure 2.2 Comment développer les objectifs du projet **19**
Figure 3.1 Théorie sous-tendant l'intervention d'un projet sur les moyens de subsistance des jeunes **29**
Figure 3.2 Éléments d'une chaîne de résultats et exemples **30**
Figure 3.3 Résultats finals volontaires *vs* résultats finals involontaires **32**
Figure 4.1 Des questions d'évaluation à la procédure d'évaluation **55**
Figure 5.1 Illustration visuelle de l'impact d'un programme **73**
Figure 5.2 Techniques expérimentales et techniques quasi expérimentales **74**
Figure 5.3 Risques inhérents à la comparaison des résultats avant et après **77**
Figure 5.4 Risques inhérents à la comparaison des participants avec des non-participants **78**
Figure 6.1 Arbre de la décision permettant de choisir une technique d'évaluation d'impact **85**
Figure 6.2 Étapes d'un tirage au sort **86**
Figure 6.3 Choisir des échantillons pour des programmes de petite ou de grande ampleur **87**
Figure 6.4 Groupe de traitement et groupe témoin dans la procédure par étapes **91**
Figure 6.5 Estimation de l'impact dans le cadre d'une promotion aléatoire **94**
Figure 6.6 Graphiques de sélection d'échantillons au moyen de l'approche par discontinuité de la régression **97**
Figure 6.7 Exemple d'analyse en double différence **100**
Figure 6.8 Appariement exact sur la base de cinq caractéristiques **103**
Figure 6.9 Barème d'admissibilité (exemple de classement selon le niveau de pauvreté) **106**
Figure 7.1 Étapes à suivre pour mener une évaluation d'impact **113**
Figure 7.2 Calendrier type pour une évaluation d'impact prospective **115**
Figure 8.1 Schéma d'une évaluation d'impact avec recours aux modèles transversaux **148**
Figure 8.2 Comparaison des coûts et bénéfices **152**

TABLEAUX

Tableau 2.1 Obstacles à court terme pour les jeunes sur le marché du travail **17**
Tableau 2.2 Liste des interventions fondées sur des données, par obstacle **22**
Tableau 3.1 Exemple de cadre logique pour un programme scolaire d'entrepreneuriat **34**
Tableau 3.2 Exemples d'indicateurs **35**
Tableau 3.3 Exemples d'indicateurs pour les projets axés sur les moyens de subsistance des jeunes **37**
Tableau 3.4 Aperçu des méthodes de collecte de données **40**
Tableau 3.5 Exemples d'hypothèses et de réponses dans le cadre du projet **43**
Tableau 3.6 Adapter les rapports aux interlocuteurs **46**
Tableau 3.7 Éléments habituels d'un budget pour le suivi **47**
Tableau 4.1 Exemples de questions d'évaluation **53**
Tableau 4.2 Le lien entre critères d'évaluation et questions d'évaluation **54**
Tableau 4.3 Compétences requises pour chaque type d'évaluation **60**
Tableau 4.4 Estimation des coûts pour les différents types d'évaluation **61**
Tableau 4.5 Aperçu des principaux types d'évaluation **64**
Tableau 6.1 Aperçu des techniques d'évaluation d'impact **107**
Tableau 7.1 Exemple de budget d'une évaluation d'impact **117**
Tableau 7.2 Impact d'évaluation – équipe et responsabilités **118**
Tableau 7.3 Aperçu des considérations éthiques à prendre en compte lors d'une étude portant sur les enfants et les jeunes **129**
Tableau 8.1 Types de questions concernant les évaluations d'impact **147**
Tableau 8.2 Estimation du coût-efficacité pour le programme *Jóvenes* **151**
Tableau 8.3 Valeur actuelle et valeur actuelle nette pour un rendement annuel de 100 dollars E.-U. **153**

AVANT-PROPOS

« Faute d'une évaluation rigoureuse, les programmes et politiques élaborés à l'intention des jeunes peuvent souffrir d'un manque de crédibilité »
— Rapport sur le développement dans le monde, 2007

Comme noté dans le Rapport de 2007 de la Banque mondiale sur le développement dans le monde, la communauté internationale ne sait toujours que très peu de choses sur les effets des interventions auprès des jeunes car seuls quelques programmes ont fait l'objet d'une évaluation d'impact. Malheureusement, cela est particulièrement vrai pour de nombreuses interventions concernant les moyens de subsistance des jeunes dans les pays en développement. Bien qu'il y ait d'immenses demandes dans le monde entier concernant les possibilités économiques pour les jeunes, nous, communauté internationale, avons été à ce jour largement incapables d'apporter suffisamment de réponses sur la marche à suivre. Certes, beaucoup d'initiatives menées par la société civile et les gouvernements permettent d'effectuer un travail remarquable. Toutefois, nous devons également reconnaître que la somme de ces interventions ne touche qu'une partie du milliard trois cents millions de jeunes du monde, dont 90 % vivent dans des pays en développement, qui ont le droit d'avoir accès aux possibilités, aux compétences, et à une deuxième chance pour gagner un revenu décent pour eux et leur famille.

Par manque de données tangibles, les gestionnaires de programme et les décideurs ne disposent que d'informations limitées sur les types de programme qui ont les meilleures retombées sur les jeunes vulnérables, les caractéristiques des programmes les plus efficaces et la meilleure manière de les mettre en œuvre. Par conséquent, dans les faits, de nombreuses interventions consacrées à l'emploi et aux moyens de subsistance des jeunes sont conçues à partir d'entretiens avec les parties prenantes, d'histoires anecdotiques ayant réussi, ainsi que des convictions et de l'expérience du concepteur du programme. Toutefois, pour que les programmes consacrés aux sources de revenu des jeunes soient couronnés de succès et que les futurs programmes soient meilleurs, il faut les ancrer sur des données solides, être bien mis en œuvre et évalués en toute rigueur, ce qui permettra aux agents d'exécution, aux décideurs et aux autres parties prenantes de tirer de nouveaux enseignements.

Le manque de données ne se traduit pas uniquement par des difficultés lors de la conception des programmes ; il complique la recherche de fonds pour des interventions axées sur les jeunes. Dans le monde entier, les dirigeants, qu'il s'agisse des autorités municipales ou nationales, reçoivent des demandes concurrentes et disposent de ressources limitées. Les maigres données sur des interventions consacrées aux moyens de subsistance des jeunes font souvent obstacle à l'allocation de ressources publiques importantes aux programmes consacrés aux jeunes car les autres acteurs présentent toujours un dossier plus convaincant pour leur domaine d'action, par exemple les investissements dans les transports ou les infrastructures énergétiques. Étant donné que les décideurs octroient des financements publics aux programmes et politiques s'appuyant sur une analyse solide des effets et du rapport coût-avantages, il nous incombe de fournir les données nécessaires prouvant qu'il est intéressant d'investir dans les possibilités économiques pour les jeunes.

Toutefois, les organismes d'exécution sont souvent préoccupés par les coûts et les efforts qu'engendrent des évaluations solides car ils pourraient consacrer leurs maigres ressources

à l'élargissement de leurs services à un nombre accru de bénéficiaires. Malheureusement, sous-estimer la valeur des évaluations traduit un manque de perspicacité. Si nous ne tirons pas systématiquement les enseignements des points forts et des points faibles de nos interventions et ne collectons pas les données nécessaires pouvant éclairer les programmes en cours et à venir, nous risquons de condamner tout le domaine de l'emploi des jeunes à un financement public insuffisant chronique.

Dans ce contexte, et pour constituer et diffuser des données sur les méthodes éprouvées permettant d'améliorer les sources de revenus des jeunes, la Banque mondiale a créé le Partenariat mondial pour l'emploi des jeunes, en 2008, qui regroupe notamment l'Institut arabe de développement urbain, la Fondation internationale pour la jeunesse, Comprendre le travail des enfants et le Réseau pour l'emploi des jeunes. Convaincu que des données plus nombreuses et de meilleure qualité peuvent améliorer la conception et les résultats des futurs programmes destinés aux jeunes et renforcer la totalité de ce domaine d'action, les membres du Partenariat défendent activement les évaluations d'impact et les produits fondés sur le savoir y afférents.

C'est dans ce contexte que nous avons élaboré le présent guide. Il offre une introduction complète et accessible à la question du suivi et de l'évaluation, ainsi qu'à son application pratique dans le domaine de l'emploi des jeunes. Nous espérons qu'il vous aidera et qu'il aidera votre organisation à décider du type d'évaluation qui vous convient étant donné les caractéristiques et le contexte de l'intervention. Si vous décidez de mener une évaluation d'impact pour votre programme, le guide décrit les instruments à disposition, les mesures concrètes à prendre et les difficultés pratiques afin de vous permettre de préparer l'évaluation.

Nous espérons sincèrement que le présent guide aidera les agents d'exécution à prendre des décisions en connaissance de cause lorsqu'ils choisiront le cadre d'évaluation qui servira leur programme et leur organisation, ainsi que l'ensemble du secteur de l'emploi des jeunes. Si, à l'avenir, nous arrivons à tirer des données plus solides à partir des travaux et d'expériences de qualité, nous aurons davantage d'arguments pour convaincre les décideurs d'adopter des méthodes ayant réussi à plus large échelle. Nous espérons que les décideurs, les acteurs du développement et toutes les autres parties prenantes continueront de travailler ensemble pour offrir aux adultes de demain les possibilités économiques qu'ils méritent.

Arup Banerji
Directeur, Protection sociale et Travail - Banque mondiale

Susana Puerto Gonzalez
Manager - Réseau pour l'emploi des jeunes

William S. Reese
Président - Fondation internationale pour la jeunesse

Furio Rosati
Directeur - Comprendre le travail des enfants

Novembre 2011

REMERCIEMENTS

Le présent guide n'aurait jamais pu être réalisé sans la contribution de nombreux collègues et amis. Nous tenons avant tout à remercier Mattias Lundberg et Wendy Cunningham pour leurs conseils d'ordre général, leurs commentaires et leur appui tout au long de la planification et de la rédaction du présent document, ainsi que le Directeur du service Protection sociale et Travail de la Banque mondiale, Arup Banerji.

Nous souhaitons également remercier nos collègues du Partenariat mondial pour l'emploi des jeunes dont les contributions substantielles nous ont permis d'enrichir diverses parties du présent guide, notamment Ibrahim Al-Turki, de l'Institut arabe du développement urbain, Awais Sufi, Travis Adkins, Dan Oliver, Jack Boyson, Susan Pezzullo, et Hannah Corey (Fondation internationale pour la jeunesse, IYF), Markus Pilgrim, Susana Puerto Gonzalez, Drew Gardiner et Rémi Ouellet (Réseau pour l'emploi des jeunes) et Gabriella Breglia (Comprendre le travail des enfants). La rédaction du présent guide a constitué un vrai travail d'équipe.

Nous sommes également extrêmement reconnaissants à nos collègues Gloria La Cava (Banque mondiale), Fiona Macaulay et Veronica Torres (Making Cents International) d'avoir relu notre travail.

Nous souhaitons saluer nos collègues de la Banque mondiale et d'ailleurs qui ont pris le temps de lire les premières ébauches du présent guide et de nous transmettre leurs précieux commentaires, suggestions, ainsi que documents de référence. Parmi eux, citons Paloma Acevedo, Juliana Arbelaez, Marinella Ariano, Carlos Asenjo Ruiz, Stefanie Brodmann, Shubha Chakravarty, Yoonyoung Cho, Viola Erdmannsdoerfer, Tanja Lohmann, Florentina Mulaj, Suleiman Namara, David Locke Newhouse, Azra Kacapor Nurkic, Yaa Oppong, Michelle Rebosio, Leopold Sarr, Haneen Ismail Sayed, Devin Silver, Stavros George Stavrou et Cornelia Tesliuc.

Nous tenons également à remercier Dino Linares (Colectivo Integral de Desarrollo), Pia Saunders (Education for Employment Foundation), Alexandre Kolev (Centre international de formation de l'Organisation internationale du travail), David Rosas (Banque interaméricaine de développement), Kristin Hausotter et Bettina Silbernagl (GIZ), Anne Golla (Centre international de recherche sur les femmes), Caroline Jenner et Shannon Wendt (Junior Achievement), Whitney Harrelson (Making Cents International), Rewa Misra (Fondation Mastercard), Leah Katerberg et Scott Ruddick (MEDA), Sonya Silva et David Woollcombe (PeaceChild International), Karen Austrian (Conseil de population), Rani Deshpande (Save the Children), Justin Sykes (Silatech), Nader Kabbani (Syria Trust for Development), Rachel Blum (USAID) et Helen Gale (Youth Business International).

Nous souhaiterions en outre féliciter Paul Gertler, Sebastian Martinez, Patrick Premand, Laura Rawlings et Christel Vermeersch pour la qualité de leur publication *Impact Evaluation in Practice*. Le présent guide s'appuie sur leurs travaux, adapte certains documents et illustrations au domaine des moyens de subsistance des jeunes et offre une présentation plus concise des méthodes d'évaluation d'impact. Nous souhaitons également remercier les participants aux cliniques d'évaluation du Réseau pour l'emploi des jeunes (YEN) à Nairobi, Beyrouth et Genève, les nombreux membres actifs du Groupsite du YEN, ainsi que les participants

à la Conférence de 2010 sur les possibilités économiques pour les jeunes dans le monde, organisée par Making Cents International, et au Congrès mondial des jeunes (2010) dont les questions et interrogations ont servi de base au présent document.

Nous remercions tout particulièrement notre coordinateur, Kris Rusch, dont les conseils, la réactivité et la propension à travailler dans des délais très serrés ont été extrêmement précieux. Nous tenons également à remercier Caroline Esclapez, Gillian McCallion et Kathryn Werthman pour leur aide à la conception du présent document.

Enfin, Kevin Hempel souhaiterait remercier Carla Rojas de sa patience.

Toute erreur contenue dans le présent document relève de notre fait.

INTRODUCTION

Objet

Les programmes visant à appuyer activement les perspectives d'emploi des jeunes existent depuis des dizaines d'années dans les pays industrialisés mais sont relativement nouveaux dans les pays en développement. De manière générale, les interventions axées sur les moyens de subsistance des jeunes soutiennent les moyens dont disposent les jeunes pour gagner leur vie et concernent notamment la formation, les services publics, l'entrepreneuriat des jeunes et les services financiers. De manière plus précise, de nombreux agents d'exécution définissent les programmes axés sur les sources de revenus des jeunes comme étant des activités ciblant principalement les groupes vulnérables et exclus de l'économie informelle, et mettant particulièrement l'accent sur l'emploi indépendant. Le présent guide adopte la définition la plus large et inclut la mise en valeur des ressources humaines dans le secteur formel.

En tant que secteur relativement neuf et novateur, peu d'interventions ont été évaluées de manière rigoureuse. En réalité, la plupart des agents d'exécution ne pourraient citer que quelques exemples. Que signifie vraiment le terme *rigueur* ? Quelles méthodes sont ou ne sont pas rigoureuses ? Pour les agents d'exécution, il peut souvent sembler évident que notre intervention aboutit aux résultats escomptés. Pourquoi dépenser nos ressources limitées dans une évaluation coûteuse au lieu d'utiliser cet argent pour fournir des services à davantage de jeunes ?

Pour ceux qui ne participent pas directement à l'intervention, son efficacité n'est pas toujours évidente. Les décideurs et les donateurs veulent des résultats crédibles et transparents répondant à certaines normes minimales fiables. Ils souhaitent souvent des évaluations basées sur les méthodes de recherche utilisées en sciences sociales qui peuvent fournir des estimations solides sur la manière dont l'intervention a touché un participant lambda du programme. Les agents d'exécution, bien qu'ils souhaitent fournir des informations de qualité sur leurs programmes, peuvent quant à eux estimer que les évaluations rigoureuses, complexes, qui peuvent s'avérer coûteuses et utiliser d'autres ressources, sont rarement réalistes et souvent inaccessibles.

Domaines d'intervention des programmes de développement de sources de revenus pour les jeunes

- *Formation pratique et professionnelle*
- *Emploi subventionné, notamment les subventions salariales ainsi que les programmes concernant les travaux publics et les services publics*
- *Services de l'emploi, notamment l'aide à la recherche d'emploi et au placement*
- *Entreprises et entrepreneuriat des jeunes*
- *Services financiers comprenant les jeunes*
- *Programmes non traditionnels pour les groupes exclus*
- *Réglementation du marché du travail touchant les jeunes*

Sources : Betcherman et al. (2007); Cunningham, Sanchez-Puerta, and Wuermli (2010); DFID (1999).

Destinataires

Le présent document est un guide d'introduction rédigé pour les agents d'exécution n'ayant que peu ou pas de connaissances sur les évaluations d'impact ou les méthodes de recherche quantitative, mais qui souhaitent néanmoins faire la preuve des résultats obtenus. Il s'adresse aux administrateurs de programmes et aux responsables du suivi et évaluation au niveau local dans toutes les organisations exerçant dans le domaine des sources de revenus des jeunes : organisations non-gouvernementales (ONG) locales et internationales, autorités locales et nationales, et donateurs bilatéraux et multilatéraux.

Étant donné les différents parcours et expériences des agents d'exécution, il est impossible d'adapter ce guide de manière égale à tout un chacun. Toutefois, nous avons essayé d'aborder de manière complète les méthodes d'évaluation des interventions afin que les lecteurs puissent s'identifier aux parties répondant au mieux à leurs intérêts et à leurs besoins.

Objectif

L'objectif du présent guide est de fournir aux lecteurs les concepts et instruments de base nécessaires pour prendre des décisions en connaissance de cause sur la meilleure façon d'évaluer leur programme. Nous souhaitons faire clairement comprendre les options qui existent en matière d'évaluation et les aspects que les agents d'exécution doivent prendre en compte au moment de choisir

celle qui leur convient, selon les objectifs d'apprentissage et le contexte opérationnel. En outre, nous décrivons comment réaliser une évaluation d'impact si elle s'avère être la méthode d'évaluation choisie.

L'objectif général est de renforcer la base sur laquelle les programmes sont élaborés et les politiques formulées en augmentant le nombre d'évaluations de la qualité dans le domaine des sources de revenus des jeunes, ce qui permettra de multiplier les interventions réussies et de les reproduire ailleurs.

But du guide

Le guide traite du suivi et de l'évaluation des interventions dans le domaine des sources de revenu des jeunes et met particulièrement l'accent sur l'évaluation d'impact. Même si les termes *suivi* et *évaluation* vont souvent de pair, ils se réfèrent à des activités assez différentes.

Le *suivi* porte sur la mise en œuvre et les progrès d'une intervention afin de soutenir l'administration du programme. Il s'agit :

- De collecter les données relatives à une mise en œuvre particulière et aux indicateurs de résultats.
- D'évaluer le respect des programmes de travail et des budgets.
- D'utiliser des informations pertinentes pour la gestion de projet et la prise de décisions.
- D'un processus continu.
- De répondre à la question suivante : « Exécutons-nous bien le projet ? »

L'évaluation fait le bilan de la conception, de la mise en œuvre ou des résultats d'une intervention afin de soutenir une nouvelle planification. Il s'agit :

- De collecter les données relatives à la conception, à la mise en œuvre et aux résultats d'un projet.
- D'examiner la pertinence, la bonne marche, l'efficacité et la durabilité d'un projet.
- De collecter des informations utiles quant aux effets de l'intervention.
- D'un processus périodique ; généralement annuel, effectué à la fin d'un projet, et comprenant la suite donnée au projet.
- De répondre à la question suivante : « Exécutons-nous le bon projet ? »

Dans l'idéal, le suivi et l'évaluation devraient faire partie intégrante du programme, dès sa conception. À vrai dire, une évaluation exacte de la réussite d'une intervention est presqu'impossible si l'évaluation n'est pas considérée comme une priorité jusqu'à ce que le programme touche à sa fin.

Une *évaluation d'impact* est un type d'évaluation qui mesure les changements apportés au bien-être des personnes, familles ou communautés concernées par une intervention en particulier. Elle répond à la question suivante : Que serait-il advenu aux bénéficiaires sans l'exécution de ce programme ? Par exemple, si un jeune diplômé d'un programme de formation professionnelle trouve un travail, s'agit-il d'un résultat direct du programme ou cette personne aurait-elle de toute façon trouvé du travail ? Comparer la situation des participants à celle d'un groupe soigneusement établi n'ayant pas participé au programme permet d'établir le lien de cause à effet. Autrement dit, une évaluation d'impact nous permet d'attribuer tout changement observé sur le bien-être des bénéficiaires du programme à l'efficacité de notre intervention.

Une évaluation d'impact est un type d'évaluation parmi d'autres, qui a ses avantages et ses limites. Nous pensons que toute intervention n'a pas nécessairement besoin d'être assortie d'une évaluation d'impact et que l'évaluation devrait appuyer l'élaboration de programmes, et non l'inverse.

[Définition]

Suivi : Processus continu de recueil et d'analyse d'informations pour voir comment un projet, un programme ou une politique est mis en œuvre et mise en regard des résultats obtenus et des résultats escomptés.

Évaluation : Diagnostic systématique et objectif de la conception, de la mise en œuvre et des résultats d'un projet en cours ou achevé pour en déterminer la pertinence, la réalisation des objectifs fixés, la bonne marche, l'efficacité, les effets et la durabilité.

Évaluation d'impact : Type d'évaluation mené pour apprécier les changements apportés au bien-être des personnes, des foyers ou des communautés qui peuvent être attribués à une intervention en particulier.

Sources : d'après Gertler et al. (2011); Kusek and Rist (2004); OECD (1991).

Toute évaluation doit être adaptée aux caractéristiques et au contexte opérationnels de l'intervention en question tout en s'inscrivant dans un cadre plus large fondé sur un système de suivi établi. Cela dit, nous pensons que nous pourrions beaucoup apprendre si nous utilisions les méthodes d'évaluation d'impact plus souvent.

Le présent guide se distingue des travaux existants de trois grandes façons:

- Premièrement, nous appliquons directement les concepts de suivi et d'évaluation – et d'évaluation d'impact en particulier – au secteur des moyens de subsistance des jeunes. L'ouvrage présente des exemples réels, des témoignages, des indicateurs et des difficultés concrètes rencontrées dans l'évaluation des interventions axées sur les sources de revenu des jeunes.
- Deuxièmement, nous cherchons l'équilibre entre les outils pratiques qui privilégient le suivi et l'évaluation de manière générale (par exemple, Gosparini et al. 2004; Kellogg 1998; Kellogg 2004) et d'autres publications spécifiquement axées sur l'évaluation d'impact (par exemple, Baker 2000; Duflo, Glennerster et Kremer 2006; Gertler et al. 2011; Khandker, Koolwal et Samad 2010; Ravallion 2008).
- Troisièmement, nous nous adressons expressément aux agents d'exécution dans le domaine des sources de revenu des jeunes qui ne connaissent pas les méthodes de recherche et d'évaluation et qui souhaitent consulter une illustration succincte, mais complète, du suivi et de l'évaluation, ainsi que de leur mise en pratique dans leur travail quotidien. Par conséquent, contrairement aux publications susmentionnées, le présent guide vise à exposer - plus brièvement et sous le prisme des jeunes- le contenu de chacune d'entre elles. Pour une présentation plus complète de l'évaluation d'impact et de sa mise en œuvre dans le contexte du développement, nous encourageons les lecteurs à consulter *L'évaluation d'impact en pratique* (Impact Evaluation in Practice), de Gertler et al. (2011).

Études de cas

Tout au long du guide, nous utilisons le projet de Fonds d'action sociale pour le Nord de l'Ouganda (NUSAF) pour illustrer les principaux points de chaque note. Nous avons choisi le NUSAF parce qu'il contient plusieurs facettes d'un programme type axé sur les sources de revenu pour les jeunes et que l'évaluation d'impact de ce projet s'est heurtée à de nombreuses difficultés. Certes, le NUSAF est relativement vaste par rapport à de nombreux autres projets en la matière. Mais, comme nous allons le voir, des évaluations d'impact peuvent être menées pour des programmes plus petits. Nous espérons que les lecteurs trouveront des points de l'étude de cas suffisamment proches de leur situation.

Étude de cas du NUSAF : Historique

Informations générales

Nom du projet :	Programme « Youth Opportunities Program » du NUSAF visant à améliorer les perspectives d'emploi des jeunes
Groupe cible :	Jeunes pauvres âgés de 15 à 35 ans dans une région ougandaise en situation d'après-conflit
Nombre de bénéficiaires :	Plus de 8 000
Budget :	1,6 million de dollars E.-U

(suite page suivante)

Étude de cas du NUSAF : Historique (suite)

Cadre du projet

Pendant vingt ans, la plupart de l'Ouganda a connu une croissance économique, une sécurité physique et une stabilité politique assorties d'une hausse des niveaux éducatif et sanitaire. Toutefois, les districts du nord du pays étaient à tous égards à la traîne. L'activité commerciale était traditionnellement située dans le sud-ouest et le centre du pays du fait d'un type de développement précolonial et colonial, de la proximité des partenaires commerciaux et de l'existence d'infrastructures.

En outre, vingt ans de guerre civile et d'insécurité dans le nord du pays (et les pays voisins) ont déstabilisé l'économie et la société de la région. La quasi-totalité des zones du nord ont connu une forme d'insécurité physique (insurrection armée, déplacement interne, vol de bétail, etc.). La guerre civile dans les districts ethniques d'Acholi, qui a déplacé toute la population rurale, soit près de deux millions de personnes, vient de se terminer. À mesure que l'urgence humanitaire diminuait, l'aide humanitaire disparaissait progressivement et l'aide nationale et internationale au développement augmentait fortement.

Le plan du Gouvernement ougandais en faveur de la paix, du redressement et du développement visait à consolider l'autorité de l'État, à reconstruire les communautés, à promouvoir la paix et la réconciliation et à revitaliser l'économie par une série de programmes, dont le NUSAF.

Activités de projet

Le Programme du NUSAF visant à améliorer les perspectives d'emploi des jeunes a ciblé les 15-35 ans vivant dans la pauvreté, et au chômage ou sous-employés. De petits groupes de jeunes s'étant organisés de manière autonome devaient définir une compétence professionnelle intéressante, identifier un institut de formation professionnelle et présenter au NUSAF du district des services techniques à financer.

Ce programme s'articulait autour de deux grands volets.

Le premier volet permettait un transfert de liquidités de 7 000 dollars E.-U. maximum aux groupes locaux de jeunes pour qu'ils s'inscrivent à l'institut de formation professionnelle, achètent le matériel de formation et s'acquittent des frais de démarrage du commerce après avoir réussi la formation.

Le deuxième volet a renforcé les capacités des ONG, des organisations communautaires et des instituts de formation professionnelle en matière de réponse aux besoins des jeunes. (La durée et l'intensité du conflit ont détruit la plupart des infrastructures du nord du pays, en particulier les établissements d'enseignement. En investissant dans ces institutions, on pourrait augmenter les capacités.)

Source : Blattman, Fiala, and Martinez (2011).

Brève description du guide et mode d'emploi

Le guide se présente comme une série de notes brèves regroupées en deux grandes parties. La première partie vise à comprendre les raisons d'une évaluation et à s'y préparer tandis que la deuxième porte sur la mise sur pied d'une évaluation d'impact. Même s'il est important de bien connaître toutes les étapes du processus, il n'est pas nécessaire de lire ce guide de A à Z. Chaque note est conçue comme un chapitre indépendant des autres, que chacun peut consulter selon ses besoins. À un lecteur qui souhaite en savoir plus sur la planification du suivi et de l'évaluation en général, nous recommandons de commencer par la première partie. En revanche, les lecteurs qui connaissent bien ce sujet mais souhaitent en savoir plus sur l'évaluation d'impact seront plus intéressés par la deuxième partie. Le tableau ci-après indique les notes les plus pertinentes pour chaque lecteur.

Guide du lecteur

PARTIE I: Jeter les bases d'une évaluation Les quatre notes de cette partie décrivent comment se préparer à une évaluation d'impact.						
Note	**Description**	**Décideurs**	**Administrateurs de programme**	**Responsables du suivi et évaluation**	**Chercheurs et personnes chargées de l'élaboration des programmes**	**Experts en évaluation d'impact**
1	Porte sur l'importance de l'évaluation et son appui aux objectifs programmatiques et organisationnels.	✓	✓			
2	Se penche sur quelques-unes des questions essentielles concernant la conception de programmes auxquelles il faudrait avoir répondu avant de passer au suivi et à l'évaluation.		✓			
3	Présente les principales étapes de l'élaboration d'un système de suivi, socle indispensable de toute évaluation.		✓	✓		
4	Pose la question du type d'évaluation qui convient le mieux à chaque programme. La réponse dépend des objectifs d'apprentissage, du contexte et des caractéristiques du projet et des ressources disponibles.	✓	✓	✓	✓	✓
PARTIE II: Renforcer les connaissances sur le programme au moyen de l'évaluation d'impact Les notes de cette partie présentent l'évaluation d'impact et donne des orientations concrètes sur sa mise en œuvre dans le cadre de l'élaboration de programmes relatifs aux moyens de subsistance des jeunes.						
Note	**Description**	**Décideurs**	**Administrateurs de programme**	**Responsables du suivi et évaluation**	**Chercheurs et personnes chargées de l'élaboration des programmes**	**Experts en évaluation d'impact**
5	Présente les principales caractéristiques d'une évaluation d'impact et explique pourquoi certaines méthodes d'évaluation couramment utilisées ne répondent pas aux mêmes critères de qualité.	✓	✓	✓	✓	
6	Se penche sur les instruments et méthodes d'évaluation d'impact et explique leur fonctionnement et ce que cela exige. Fournit également un arbre de décision pour aider le lecteur à réfléchir à la méthode la mieux adaptée à son cas.		✓	✓	✓	
7	Passe du niveau conceptuel au niveau pratique en exposant les principales mesures à prendre pour mener une évaluation d'impact et fournir des ressources pratiques. Ces mesures couvrent la totalité du processus, des premiers préparatifs à la communication des résultats.		✓	✓	✓	✓
8	Présente les instruments nécessaires pour accroître la pertinence des évaluations d'impact. Contient un aperçu de la variété des questions sur les évaluations d'impact, de l'utilisation de méthodes mixtes et des analyses coût-efficacité et coût-avantage.				✓	✓

PARTIE I

JETER LES BASES D'UNE ÉVALUATION

NOTE 1 : Pourquoi évaluer ?

On réussit si l'on sait ce qui fonctionne bien.

— Bill Gates

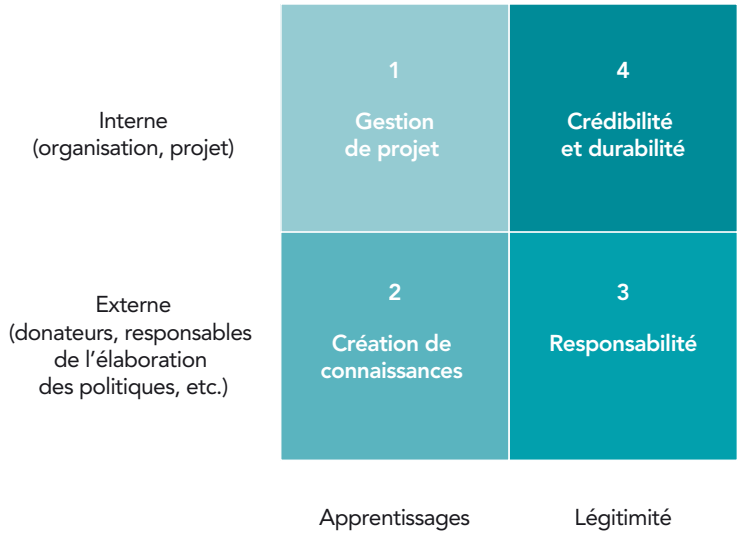
Les interventions axées sur les sources de revenus des jeunes sont complexes en soi. Comme nous visons un groupe cible dynamique, en transition au niveau biologique, social et juridique, les interventions que nous mettons en place sont très diverses et ont des résultats dans un large éventail de secteurs. Une évaluation adéquate de ces interventions, certes difficile, est un ingrédient essentiel du succès.

L'objectif de la présente note est de fournir un aperçu de la manière dont les organisations prises individuellement et l'ensemble du domaine bénéficient des connaissances tirées d'une évaluation formelle, en particulier d'une évaluation d'impact. D'après nous, une évaluation a deux grands objectifs : l'apprentissage et la légitimation. Pour chacun de ces objectifs, il y a les cibles interne et externe (voir figure 1.1). Il y a en tout quatre bonnes raisons de mener des évaluations :

- Pour administrer des projets ;
- Pour dégager des connaissances ;
- Pour veiller au respect de l'obligation de rendre des comptes ; et
- Pour renforcer la crédibilité et la durabilité de notre organisation.

Ces quatre raisons sont présentées ci-après.

FIGURE 1.1 Avantages de l'évaluation



Gestion de projets

Les évaluations nous permettent de mesurer la valeur réelle de notre travail. La plupart d'entre nous souhaite savoir ce que nos programmes changent dans la vie des jeunes pour lesquels nous travaillons. Notre projet a-t-il atteint les résultats escomptés ? Qui en a tiré le plus/le moins partie ? Les évaluations permettent de répondre à ces questions, entre autres questions, en évaluant la pertinence, l'efficience, l'efficacité, l'impact et la durabilité d'une intervention.

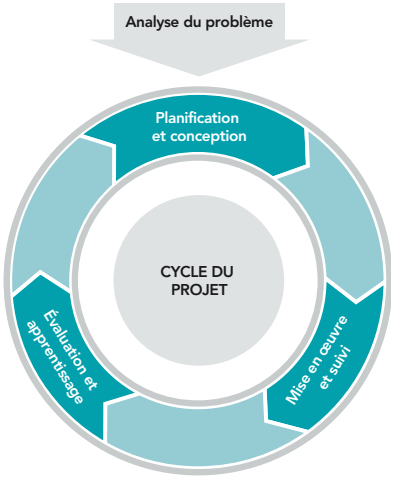
Les évaluations favorisent l'apprentissage. Grâce à l'évaluation de la conception, de la mise en œuvre ou des résultats d'une intervention, les évaluations améliorent les connaissances organisationnelles. Elles nous permettent de déterminer les volets de notre intervention qui ont réussi et ceux qui n'ont pas apporté les résultats escomptés. Par exemple, une évaluation peut révéler que la formation offerte pendant une intervention était insuffisante, ce qui nous apprend peu de choses, ou, au contraire, que la formation était trop intensive, qu'elle submergeait les élèves et qu'ils ont abandonné. De même, une évaluation peut nous aider à comprendre les conséquences imprévues de notre projet, par exemple une augmentation de la consommation d'alcool des parents liée au fait de fournir aux filles des possibilités de revenus.

Les évaluations soutiennent une nouvelle planification. Les évaluations fournissent aux gestionnaires de programmes les informations nécessaires pour prendre des décisions stratégiques sur les changements nécessaires en matière de conception, de planification et de mise en œuvre des projets. Même si les évaluations en général (et les évaluations d'impact en particulier) fournissent des informations de manière périodique et non continue, elles constituent un volet précieux du cycle du projet. Même les évaluations rétrospectives sont généralement tournées vers l'élaboration des prochains programmes (UNICEF 1991). Comme le montre la figure 1.2, l'évaluation applique les enseignements tirés des interventions en cours ou achevées à la planification et à la conception des futurs programmes. Une évaluation bien conçue aide les agents d'exécution à faire les coupes budgétaires nécessaires aux programmes en faveur des jeunes qui ne réalisent pas leurs objectifs tout en soutenant les programmes qui atteignent ou pourraient atteindre de bons résultats. Sans données issues d'une bonne évaluation, on risque davantage de tirer des conclusions erronées quant à la poursuite d'un programme et aux ressources à y allouer (Banque mondiale 2009).

« Disposer de bonnes données sur les raisons pour lesquelles les jeunes abandonnaient la formation nous ont permis de justifier l'attribution d'allocations. Cela nous a permis de faire chuter le taux d'abandon de 35 à 9 %. »

— Administrateur de programmes, ONG antillaise

FIGURE 1.2 Le cycle du projet



Création de connaissances

Le domaine des sources de revenus des jeunes se caractérise par la grande absence de données solides. Même si nos institutions font bien de collecter régulièrement des données pour assurer le suivi et évaluer les résultats de manière ordinaire (par exemple en menant des comparaisons simples avant/après et par le biais de groupes de discussion), nous sommes souvent incapables de constituer des connaissances généralisables dont tout le domaine pourrait tirer parti (Savedoff, Levine et Birdsall 2006). L'acquisition de ces connaissances exige généralement de réaliser des évaluations d'impact utilisant des méthodologies spécifiques pour fournir des estimations fiables quant à la réussite d'une intervention donnée.

Malgré les milliards de dollars consacrés à l'exécution de programmes axés sur les moyens de subsistance des jeunes, il n'existe que peu d'évaluations d'impact. Par exemple, dans une étude mondiale sur les données relatives aux interventions en faveur de l'emploi des jeunes, Betcherman et al. (2007) n'ont relevé que trois évaluations de qualité de programmes d'entrepreneuriat des jeunes. De même, on ne sait que peu de choses sur les stratégies de promotion d'autres moyens de subsistance, par exemple l'école de la deuxième chance, les travaux publics programmés, l'éducation financière des jeunes et les services financiers aux jeunes (voir encadré 1.1). Même si, ces dernières années, des efforts accrus ont été déployés établir des données solides, davantage de connaissances sont nécessaires.

ENCADRÉ 1.1 Données existantes sur l'emploi des jeunes

Betcherman et al. (2007) ont mené une étude mondiale sur les interventions dans le domaine de l'emploi des jeunes et ont conclu que « seul un programme sur dix [était] évalué de manière à mesurer son impact net et son coût. » Les types d'interventions où les lacunes sont les plus graves sont les programmes d'emploi subventionné, et les interventions liées à l'entrepreneuriat des jeunes, aux services d'emploi et aux réformes réglementaires. Au niveau régional, les données sont particulièrement rares en Asie, au Moyen-Orient et en Afrique du Nord, ainsi qu'en Afrique subsaharienne. Une base de données actualisée sur les interventions dans le domaine de l'emploi des jeunes se trouve sur le site Internet de l'inventaire sur l'emploi des jeunes (Youth Employment Inventory, http://www.youth-employment-inventory.org).

Card, Kluve et Weber (2009) ont mené une méta-analyse des programmes du marché du travail dans les pays de l'OCDE. En comparant les types de programmes, ils ont établi que les programmes d'emploi subventionné dans le secteur public avaient le plus faible impact. Les programmes d'aide à la recherche d'emploi ont un impact relativement favorable à court terme tandis que les programmes de formation présentielle et en cours d'emploi obtiennent davantage d'effets à moyen terme. Les auteurs ont conclu que l'impact des programmes pour les jeunes dans les pays de l'OCDE pouvait être moins positif que celui de programmes non ciblés.

Ibarrarán et Rosas Shady (2009) ont résumé les conclusions d'évaluations rigoureuses de programmes de formation professionnelle en Amérique latine. Contrairement aux données obtenues dans les pays développés, les conclusions suggèrent que ces programmes ont des effets positifs sur l'emploi et la qualité de l'emploi des personnes suivant la formation, en particulier les femmes et les participants les plus jeunes. Les auteurs admettent qu'il n'y a toujours aussi peu de connaissances en matière d'effets de ces interventions sur le long terme en Amérique latine.

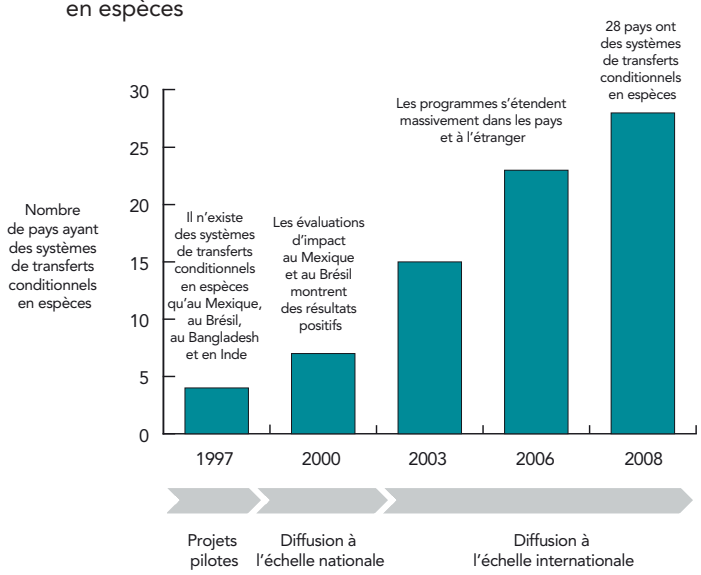
Cunningham, Sanchez-Puerta et Wuermli (2010) résument les données actuelles sur les programmes concernant le marché du travail pour les jeunes, classés par la contrainte à laquelle ils essaient de s'attaquer. Tout en plaidant pour un apprentissage et une évaluation rigoureux pour tous les types de programmes, les auteurs ont trouvé que les lacunes étaient particulièrement graves dans les interventions telles que l'école de la deuxième chance, la formation aux compétences comportementales et à l'entrepreneuriat, les programmes de travaux publics et de services publics, l'aide à la recherche d'emploi grâce aux moyens technologiques, la certification des compétences et la micro-finance, entre autres.

Note : Veuillez vous reporter à la section ressources du présent guide pour consulter une liste des évaluations d'impact terminées et en cours dans le domaine des moyens de subsistance des jeunes.

Le manque d'études rigoureuses, malgré une forte demande, restreint fortement les investissements à grande échelle dans le secteur. L'absence de données limite également l'obtention de l'aide des autorités publiques pour les interventions en faveur des moyens de subsistance des jeunes. Les responsables des services de l'État veulent généralement connaître des estimations quant à l'impact et au rapport coût-avantages avant d'investir dans de grands programmes. L'absence de données en la matière empêche souvent de défendre convenablement nos programmes par rapport à d'autres interventions, par exemple le développement d'infrastructures, où il y a beaucoup plus de données. Améliorer la base de connaissances faciliterait donc l'intensification et la réédition.

Cette possibilité est devenue évidente dans d'autres domaines. L'augmentation des données sur les programmes de transferts conditionnels en espèces, par exemple, ont permis à la communauté internationale de favoriser des interventions à grande échelle dans ce domaine dans le monde entier. Le Mexique et le Brésil n'étaient que deux des quatre pays au monde dotés de programmes de transferts conditionnels en espèces en 1997. Les évaluations d'impact ont permis d'élargir le modèle à 28 pays en 2008 (voir figure 1.3).

FIGURE 1.3 Des données aux politiques : programmes de transferts conditionnels en espèces


28 pays ont des systèmes de transferts conditionnels en espèces
Les programmes s'étendent massivement dans les pays et à l'étranger
Nombre de pays ayant des systèmes de transferts conditionnels en espèces
30
25
20
15
10
5
0
Il n'existe des systèmes de transferts conditionnels en espèces qu'au Mexique, au Brésil, au Bangladesh et en Inde
Les évaluations d'impact au Mexique et au Brésil montrent des résultats positifs
1997
2000
2003
2006
2008
Projets pilotes
Diffusion à l'échelle nationale
Diffusion à l'échelle internationale

Sources : Fiszbein et Schady (2009) ; Rawlings et Rubio (2005).

Constituer systématiquement un socle de connaissances sur ce qui marche en matière de renforcement des possibilités économiques des jeunes permettrait d'améliorer l'efficacité et la bonne marche de nos travaux en apportant des connaissances essentielles à l'activité des agents d'exécution et des décideurs, et renforcerait à terme le domaine dans son ensemble.

Obligation de rendre des comptes

Outre l'amélioration de l'apprentissage en interne et dans tout le secteur, les évaluations renforcent la légitimité de nos opérations. Les organismes de financement et la société exigent de plus en plus aux programmes de développement de rendre des comptes. Par conséquent, les évaluations - en particulier les évaluations d'impact - peuvent fournir des données nécessaires, solides et crédibles, sur l'obtention des résultats escomptés par un programme particulier.

Nos ressources ne devraient pas être considérées comme acquises. Dans la quasi-totalité

L'initiative en matière de formation professionnelle bénéficiant de financements publics en Amérique latine, les programmes Jóvenes, ont associé formation technique et formation générale, stages et autres services. Les évaluations ont montré leurs effets mesurables, par exemple une augmentation des taux d'emploi et des salaires qui touchaient, dans certains pays, plus de 10 % des jeunes et des filles. Les programmes Jóvenes ont donc été rapidement réédités sur tout le continent. Les nouveaux programmes étant également évalués, on a pu montrer la diversité des effets selon les pays et les contextes et souligner que de nombreuses évaluations étaient nécessaires avant d'arriver à une généralisation des résultats. Bien que l'expérience des programmes Jóvenes demeure une exception dans le secteur des moyens de subsistance des jeunes, en ce qui concerne l'évaluation systématique des effets, elle montre que nous ne pouvons élargir ce domaine qu'en diffusant de meilleures connaissances et données tirées des centaines d'interventions que nous exécutons chaque année.

des cas, nos projets sont financés par des fonds publics ou privés, tels l'aide publique au développement (essentiellement l'argent des contribuables) ou des dons privés. Dans les deux cas, quelqu'un nous a confié, directement ou indirectement, la charge d'utiliser son argent du mieux possible pour améliorer la vie des jeunes. Le fait que l'on nous confie l'élaboration et la mise en œuvre d'un projet consacré aux moyens de subsistance des jeunes signifie que cet argent ne va pas servir à construire des routes rurales, à améliorer un projet de prévention du VIH/sida ou à acheter des fournitures scolaires. Étant donné la rareté des ressources, il semble naturel d'utiliser les évaluations pour rendre compte en toute honnêteté de nos travaux : comment l'argent a été utilisé, les activités financées et les résultats obtenus.

Il nous incombe de garantir la meilleure utilisation possible des fonds. Les interventions en matière de développement sont intrinsèquement complexes et il serait illusoire de s'attendre à un taux de réussite de 100 %. Par ailleurs, un projet n'améliore pas le bien-être des personnes uniquement parce qu'il manifeste de bonnes intentions. Pour veiller à ce qu'un programme donné fasse plus de bien que de mal et garantir que les avantages de l'investissement soient supérieurs aux avantages qui seraient obtenus si on utilisait autrement les ressources, les agents d'exécution devraient toujours accorder rang de priorité à une évaluation soigneuse de l'efficacité de l'intervention (Jones et al. 2009).

Les donateurs internationaux se penchent de plus en plus sur des études impact rigoureuses mesurant la réussite des programmes qu'ils financent. En 2011, le Ministère britannique du développement international (DFID) et l'Agence des États-Unis pour le développement international (USAID) ont davantage mis l'accent sur les résultats. Voir

***DFID's Business Plan 2011–2015, Section 2** « Make British aid more effective by improving transparency and value for money »*
http://www.dfid.gov.uk/Documents/DFID-business-plan.pdf

USAID's Evaluation Policy 2011
http://www.usaid.gov/evaluation/USAIDEvaluationPolicy.pdf

Crédibilité et durabilité

Les évaluations contribuent à accroître la légitimité du projet et à renforcer la réputation de l'organisme d'exécution. Cet argument est rarement mentionné dans les travaux existants, mais, dans les faits, il s'agit peut-être de l'une des principales raisons encourageant à mener une évaluation d'impact car elle sert directement les intérêts du programme et de l'organisme d'exécution.

Les évaluations d'impact peuvent renforcer la crédibilité et la réputation de notre organisation. La rareté des évaluations de qualité fait qu'on leur octroie une attention particulière. Par conséquent, le simple fait qu'une organisation ou un projet accepte de mener une évaluation d'impact traduit déjà la qualité des normes en place lors de l'élaboration du programme. Si l'étude montre de bons résultats, les retombées pour l'organisation et le programme vont être très grandes. Imaginez que parmi les centaines d'acteurs du domaine, *vous* soyez celui qui peut prouver que sa méthode marche, que son programme réussit à fournir des possibilités de revenus aux jeunes hommes et aux jeunes femmes clairement supérieures à celles qu'ils auraient obtenues s'ils n'avaient pas participé au programme. La différence est que vous n'êtes pas seulement capable de *prétendre* que votre intervention est efficace, vous pouvez le *prouver*. Cela fait une grande différence aux yeux des donateurs et des décideurs qui, avant l'évaluation, ne pouvaient faire la différence entre les effets de votre intervention et les effets allégués par de nombreux autres administrateurs de programmes.

Pouvoir sortir du lot peut avoir plusieurs effets positifs sur le projet et l'organisation. Une évaluation positive peut aider, lors du plaidoyer et de la levée de fonds, à obtenir un appui plus important des donateurs, des gouvernements et de la population. Avec un appui de l'État et un soutien politique accru, notre projet et notre organisation peuvent rapidement devenir la référence dans le domaine, ce qui, à son tour, entraîne souvent une augmentation de la demande de services. Nous pouvons ainsi nous attendre à étendre nos services dans le pays et à l'étranger.

Prenons, par exemple, le cas du *Colectivo Integral de Desarrollo* au Pérou. En 2003, l'organisation était l'une des toutes premières à fournir des données rigoureuses montrant que leur modèle de promotion d'entrepreneurs jeunes à faible revenu augmentait la taille des activités, en améliorait la survie et dopait les revenus. Grâce à ces données, le *Colectivo Integral de Desarrollo* a reçu de nombreuses récompenses et n'a plus eu de difficultés à assurer le financement de ses programmes. Il reçoit désormais l'appui d'un programme de subvention de la Banque interaméricaine de développement et étend son modèle à l'Amérique centrale et aux Caraïbes (voir encadré 1.2.).

« Comme nous pouvions prouver comment l'emploi des jeunes s'était amélioré, le gouvernement nous a invité à participer à la conception d'un programme sur l'employabilité sous l'égide du nouveau Président. »

— Directeur de programme d'une ONG chilienne

ENCADRÉ 1.2 Avantages d'une évaluation d'impact

Réponse du président de l'ONG péruvienne *Colectivo Integral de Desarrollo* à la question « Quel parti pensez-vous que votre organisation a tiré de la tenue d'une évaluation d'impact ? »

« Elle a amélioré la qualité de notre intervention. »

« Elle a amélioré la crédibilité du programme. »

« Elle a amélioré la valeur de notre marque aux yeux des donateurs. »

« Nos services ont été davantage demandés. »

« Nous avons obtenu une reconnaissance nationale et internationale. »

« Nous sommes une institution qui sert de modèle dans d'autres contextes et d'autres pays. »

Source : Dino Linares, Président du *Colectivo Integral de Desarrollo*, le 28 janvier 2011.

Les résultats négatifs d'une évaluation ne sont pas nécessairement mauvais

Bien que nous puissions craindre que les résultats négatifs d'une évaluation fassent que les donateurs réduisent leur financement, des évaluations qui ne confirment pas les résultats positifs d'une intervention peuvent être utiles.

Les résultats négatifs sont inévitables dans les programmes novateurs. L'innovation et la créativité sont essentielles pour aider les jeunes à réussir leur transition vers le travail. L'innovation entraîne par définition des échecs. Comme dans d'autres domaines tels que la médecine, la chimie ou la physique, l'élaboration de produits et de services couronnés de succès passent par une mise à l'épreuve, des prototypes, un affinage et une adaptation aux circonstances locales. Les échecs sont une étape nécessaire pour parvenir à un programme de pointe.

Des résultats négatifs peuvent aider à améliorer les opérations. Si, dès le début, nous pouvons comprendre les problèmes qui pourraient rendre notre intervention moins efficace, nous pourrons construire des projets couronnés de succès sur le long terme. Les résultats négatifs d'une évaluation nous montrent comment mieux élaborer un programme.

Traiter en amont des résultats négatifs renforce la crédibilité. Aucun donateur ni décideur ne s'attend à ce que chaque projet soit une grande réussite ni ne croit que cela sera le cas. Diffuser les conclusions, favorables ou non, montre que nous sommes capables de porter un regard critique sur nos actions et fait la preuve de notre volonté d'apprendre et de fonder l'élaboration de nos programmes sur des données tangibles. Certes, la pression quant à l'obtention de résultats et à la justification des budgets nous pousse à faire passer les conclusions positives avant les négatives. Mais, à long terme, une discussion honnête sur ce qui a réussi et sur ce qui a échoué aura probablement de meilleures retombées.

Un programme mené au Moyen-Orient a offert une démarche novatrice en matière de formation de jeunes femmes au poste d'assistante exécutive de femmes entrepreneures. Le programme a tiré parti de liens importants avec le secteur commercial et en particulier les femmes entrepreneures qui souhaitaient appuyer des jeunes femmes défavorisées et les rendre autonomes. Toutefois, la première évaluation a montré qu'un nombre relativement élevé de jeunes femmes abandonnaient peu après leur placement. Une enquête plus poussée sur la cause de ces abandons a montré que les femmes entrepreneures avaient des attentes très élevées mais qu'elles ne les aidaient pas suffisamment à réussir dans des environnements professionnels exigeants. Grâce à un dialogue ouvert avec le partenaire d'exécution local, le programme a revu les modules de compétences pratiques pour mieux préparer les jeunes femmes aux exigences de ces emplois et a reformulé les orientations à destination des chefs d'entreprise pour qu'elles veillent à fournir un encadrement suffisant aux nouvelles employées. L'organisme d'exécution a pu obtenir de meilleurs résultats grâce à ces correctifs à mi-parcours de la stratégie du programme, a communiqué ce qu'il avait appris aux autres donateurs et a ensuite garanti un financement supplémentaire pour étendre le programme.

Points essentiels

1. Les évaluations servent avant tout à tirer des enseignements pour le bien de notre projet et de notre organisation. Elles nous permettent de faire la preuve de la valeur réelle de notre travail et servent à concevoir et à planifier d'autres interventions.
2. Les évaluations créent une base de données plus que nécessaire au domaine des moyens de subsistance des jeunes. Des connaissances plus nombreuses et de meilleure qualité sur ce qui fonctionne et ce qui ne fonctionne pas aideront les agents d'exécution à concevoir des

interventions efficaces et à convaincre les décideurs d'y apporter un financement public.

3. Les évaluations confèrent une légitimité car elles nous rendent responsables devant les donateurs et le public. Les évaluations garantissent le bon usage de l'argent des contribuables et des dons.
4. Les évaluations renforcent notre crédibilité et notre réputation. Dans un secteur où il y a peu de données solides, mener des évaluations peut avoir des retombées importantes car cela peut doper la demande de services, renforcer la marque de notre organisation et enfin garantir un appui financier durable.
5. Les évaluations n'ont pas besoin de montrer de bons résultats pour être utiles. Au contraire, on apprend de ses échecs. Résoudre en amont les résultats négatifs d'une évaluation et en parler renforcent sûrement notre crédibilité et notre réputation.

La force de la mesure des résultats

- *Si vous ne mesurez pas les résultats, vous ne pouvez pas faire la différence entre succès et échec.*
- *Si vous ne pouvez pas voir les succès, vous ne pouvez pas en tirer d'enseignements.*
- *Si vous ne pouvez pas voir les échecs, vous ne pouvez pas rectifier le tir.*
- *Si vous pouvez faire la preuve de vos résultats, vous pouvez obtenir appui et financement publics.*

Source : d'après Osborn et Gaebler (1992).

Étude de cas du NUSAF : pourquoi évaluer ?

L'évaluation d'impact du NUSAF a été engagée par le Gouvernement ougandais avec l'appui de la Banque mondiale. La principale raison de cette évaluation était l'amélioration de l'administration du programme. En estimant les effets du Programme « Youth Opportunities Program » visant à améliorer les perspectives d'emploi des jeunes sur les moyens de subsistance et le bien-être des jeunes dans le nord du pays, l'étude a tenté d'éclairer les futurs cycles de programme et son éventuelle intensification.

L'évaluation d'impact devait aussi combler une lacune importante quant à la compréhension de l'efficacité des programmes de formation aux compétences professionnelles et d'entrepreneur, en particulier dans le contexte africain. En allouant des subventions à la formation professionnelle et à la constitution d'un capital de démarrage pour mettre sur pied des entreprises productives, le Programme réunit les principales composantes des deux types de programmes d'emploi les plus courants. Étant donné que l'on en sait peu sur l'efficacité de cette démarche, l'évaluation permettra d'éclairer le domaine des moyens de subsistance des jeunes.

Source : Blattman, Martinez et Fiala (2009).

À lire

Savedoff , W., Levine, R., and Birdsall, N. 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation.* Washington, DC: Center for Global Development. http://www.cgdev.org/content/publications/detail/7973.

Notes

NOTE 2 : Examen de la conception d'un projet

Mon tailleur est le seul homme à faire preuve d'intelligence.
Il reprend mes mesures chaque fois qu'il me voit
tandis que les autres utilisent les anciennes
et s'attendent à ce qu'elles me correspondent toujours.

— George Bernard Shaw

Imaginez-vous dans les situations suivantes :

Le maire de Tripoli (Liban) demande à votre organisation une aide technique pour lutter contre le chômage des jeunes dans la ville. Que faire ?

L'administrateur du bureau dans votre pays, le Rwanda, est intéressé par les programmes d'auto-emploi des jeunes. Quelle marche à suivre recommandez-vous ?

Votre plan stratégique pour les trois prochaines années met davantage l'accent sur la transition des jeunes vers le travail. Dans quels programmes d'emploi des jeunes devriez-vous investir ?

Les administrateurs de programmes doivent prendre des décisions difficiles sur ces sujets et sur d'autres questions relatives à l'élaboration de programmes. Pour trouver les solutions adéquates, nous devons comprendre le contexte spécifique et concevoir un programme solide. Par conséquent, avant de concocter un système de suivi et d'évaluation, nous devons veiller à ce que notre intervention soit soigneusement planifiée : avons-nous les connaissances adéquates sur les besoins des personnes que nous essayons d'aider ? Comprenons-nous pourquoi existent certaines situations telles que le chômage des jeunes et l'exclusion sociale ? Avons-nous un objectif clair que nous voulons atteindre ? Et utilisons-nous les expériences et les données existantes lorsque nous élaborons notre intervention en vue de l'atteindre ?

Analyse du problème : comprenons-nous le groupe cible et le contexte local ?

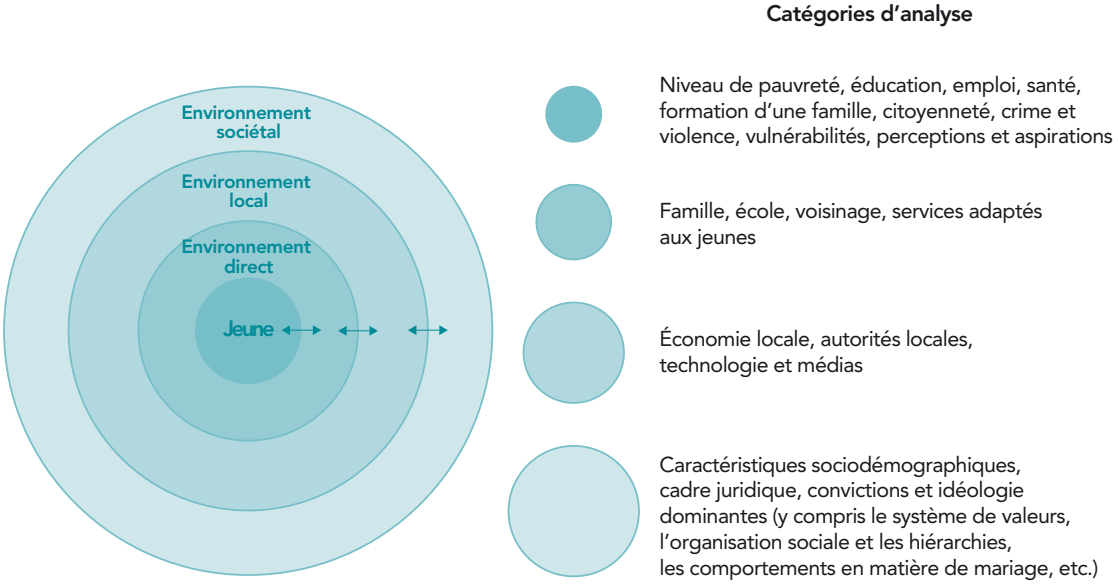
Lorsque nous décidons de mener une intervention pour aider les jeunes, nous partons du principe qu'il y a un écart entre le *statu quo* et ce qui pourrait ou devrait être fait. Cet écart entre la situation actuelle et la situation souhaitée est ce que nous appelons généralement un *besoin*. Nous devons comprendre ces besoins avant de commencer à intervenir.

Utiliser les évaluations transversales relatives aux jeunes pour comprendre notre groupe cible

Les possibilités économiques peuvent rarement être comprises de manière isolée. Elles sont très souvent reliées à d'autres aspects de la vie. Par exemple, l'emploi, le chômage et le revenu peuvent déterminer la possibilité pour un individu de se marier ou de fonder une famille. Par conséquent, pour comprendre les problèmes et les besoins des jeunes, il faut presque toujours évaluer de manière approfondie les conditions de vie des jeunes, notamment leur statut socio-économique, leurs comportements et attitudes, ainsi que leurs objectifs et aspirations. Ces évaluations doivent aussi saisir d'autres facteurs importants pertinents pour le passage des jeunes à l'âge adulte, tels que la santé, la constitution d'une famille et la citoyenneté (pour plus d'informations sur le passage à l'âge adulte, veuillez vous reporter au document de la Banque mondiale de 2006).

En outre, puisque les jeunes sont influencés par divers facteurs autour d'eux, notamment la famille, les pairs, les communautés, les institutions locales et nationales, ainsi que les normes sociales, de bonnes évaluations sur les jeunes doivent également analyser l'environnement direct, local et sociétal des jeunes (voir Bronfenbrenner 1979). Une évaluation exhaustive brossera un tableau complet des besoins des jeunes et des difficultés auxquelles ils sont confrontés, ce qui nous permettra de mieux adapter notre intervention aux réalités locales (voir figure 2.1.). Les indicateurs types d'évaluation concernant les jeunes sont énumérés à l'appendice 1.

FIGURE 2.1 Environnements des jeunes


Catégories d'analyse
Environnement sociétal
Environnement local
Environnement direct
Jeune
Niveau de pauvreté, éducation, emploi, santé, formation d'une famille, citoyenneté, crime et violence, vulnérabilités, perceptions et aspirations
Famille, école, voisinage, services adaptés aux jeunes
Économie locale, autorités locales, technologie et médias
Caractéristiques sociodémographiques, cadre juridique, convictions et idéologie dominantes (y compris le système de valeurs, l'organisation sociale et les hiérarchies, les comportements en matière de mariage, etc.)

L'analyse des environnements personnel et social des jeunes grâce à des entretiens systématiques, des groupes de discussions et des observations nous aidera à déterminer le principal problème auquel nous souhaitons nous attaquer, par exemple le sous-emploi ou le chômage, ou encore l'insuffisance de l'accès aux services financiers. Nous pouvons également réaliser que des opportunités économiques limitées ne sont que l'un des nombreux problèmes que les jeunes rencontrent dans un lieu donné, ce qui peut suggérer comment établir ou adapter notre intervention pour qu'elle soit axée sur plus d'un problème.

L'évaluation est aussi importante ; elle nous aidera à déterminer notre groupe cible. Sommes-nous intéressés par tous les jeunes ou uniquement par ceux qui ont quitté le système scolaire ? Sur quelle tranche d'âge souhaitons-nous mettre l'accent ? Y a-t-il des considérations ethniques ou liées au genre que nous voudrions rendre prioritaires ? Quelles zones géographiques allons-nous cibler ? Étant donné que nous disposons de ressources limitées, nous sommes rarement en mesure de nous adresser à chaque jeune. Une évaluation transversale des jeunes peut nous aider à établir des priorités, par exemple en déterminant les groupes particulièrement vulnérables, tels ceux qui ont abandonné l'école, les jeunes filles ou les enfants des rues.

En menant une évaluation communautaire rapide du profil socio-économique des jeunes ciblés en Jordanie, la Fondation internationale pour la jeunesse a découvert que le niveau de participation des jeunes aux activités civiques dans 12 communautés ciblées était extrêmement faible (moins de 4 %). En outre, les données de l'étude ont révélé que peu d'institutions offraient des possibilités de bénévolat. Lorsque les jeunes participaient à des projets de services à la communauté, ils le faisaient principalement par leur école. Les échanges au sein des groupes de discussion ont également montré que les jeunes étaient prêts à faire du bénévolat si des possibilités surgissaient, même en l'absence de culture du volontariat dans ces communautés.

Ces conclusions ont permis d'éclairer la conception de projets de formation à certains services pour les jeunes ayant abandonné l'école et sans emploi. Elles ont aussi justifié l'allocation de subventions pour que les jeunes mènent de petites initiatives communautaires, rendant ainsi les possibilités de participation civique plus visibles et accessibles, et garantissant l'attractivité de ces possibilités auprès des jeunes. En outre, la Fondation internationale pour la jeunesse a formé le personnel d'organisations travaillant auprès des jeunes à l'efficacité du développement, à la gestion de la participation communautaire et à l'administration de programmes de bénévolat, ce qui les a aidés à mieux faire participer les jeunes à la vie de leur communauté.

Utiliser les analyses de marché pour comprendre l'économie locale

Étant donné que nous mettons l'accent sur la construction ou le renforcement des moyens de subsistance, un composant essentiel de l'analyse du contexte est l'analyse de l'économie locale. Cette dernière inclut généralement l'évaluation du marché du travail local et du marché des biens et services.

Évaluation du marché du travail local

Les évaluations d'une marché du travail visent à comprendre les modèles et tendances en matière d'emploi dans l'économie locale. Les facteurs courants à analyser sont les points suivants (Banque asiatique de développement 2007, p. 162–166) :

- **Demande de main-d'œuvre.** Conditions économiques générales ; taille des secteurs formel et informel ; secteurs ou industries dynamiques et zones géographiques qui ont besoin de main-d'œuvre ; tendances et projections dans l'industrie ; nombre d'emplois qui devraient être créés ; compétences requises par profession ; niveaux de salaire et bénéfices ; conditions de travail ; pratiques en matière de recrutement ; mentalité de l'employeur ; obstacles à l'emploi fondés sur le genre, l'âge, l'appartenance ethnique, le statut social, la religion ou d'autres raisons ; etc.
- **Offre de main-d'œuvre.** Taille et structure de la population en âge de travailler ; emploi, sous-emploi et chômage par genre, âge, niveau d'éducation, zones urbaine/rurale, secteur économique, profession, et secteurs formel/informel et public/privé.
- **Environnement institutionnel et politique.** Programmes, politiques, lois et institutions existants relatifs au marché du travail, par exemple les réglementations quant au salaire minimum, les lois relatives à la protection de l'emploi, la syndicalisation, les prestations chômage et tout autre disposition similaire. Les priorités économiques définies par secteur au niveau national, régional et local sont également intéressantes.

Évaluation du marché des biens et des services

L'évaluation du marché des biens et des services aide à déterminer si les petits producteurs peuvent participer à des activités économiques durables et comment répartir les rôles (par exemple pour les jeunes et les femmes) sur ces marchés (Penrose-Buckley 2007). Les caractéristiques communes des marchés à analyser sont :

L'initiative LEGACY du Comité international de secours au Libéria met l'accent sur le renforcement des programmes d'éducation des communautés et l'élargissement des possibilités de formation professionnelle axée sur le marché pour les femmes et les jeunes généralement exclus. Pour garantir que le programme était correctement axé sur le marché, le Comité a mené deux évaluations.

Une étude sur le marché du travail a défini les compétences monnayables qui pourraient constituer des possibilités d'emploi pour les jeunes. Un questionnaire de dix pages, associant questions à choix multiple et questions à réponse libre a permis de mieux comprendre les besoins actuels et futurs des employeurs en matière de recrutement.

En outre, le Comité a conduit une brève étude de marché pour aider deux centres de formation professionnelle à déterminer les produits et services nécessaires à d'éventuelles activités scolaires. Cette étude s'est composée de questionnaires adressés aux détaillants, aux usagers et aux fournisseurs des marchés locaux.

Ces évaluations ont notamment aidé le Comité à comprendre les besoins des filles par rapport à ceux des garçons, contribuant ainsi à accroître le nombre de filles dans les centres de formation.

Source : Beauvy-Sany et al. (2009).

- **Besoins des marchés et chaînes de valeur.** Lacunes existantes et futures quant aux produits et services pour les consommateurs ; demande de produits de base, de produits transformés et de produits semi-finis par les détaillants, les grossistes ou les entreprises de transformation ; définition des marchés local, régional et d'exportation ; détermination des acteurs actuels du marché ; et autres facteurs.
- **Stabilité du marché.** Vulnérabilité du marché aux chocs, aux saisons et aux changements de tendance ; restrictions éventuelles à l'accès au marché, et déplacements de personnes et de produits du fait d'un conflit et de l'insécurité.
- **Prix du marché.** Volatilité du prix des produits finis et des fournitures ; effets sur les prix d'un nombre accru de producteurs ; inflation ; coûts de transaction.

Les études de marché consistent généralement à combiner analyse de données existantes et études sur les employeurs ou les détenteurs de petites affaires. Les personnes interrogées peuvent fournir des indications importantes sur les perspectives d'emploi dans des secteurs particuliers, la manière dont les décisions relatives au recrutement sont prises, les principales contraintes qui pèsent sur les entreprises formelles et informelles, leur opinion sur les jeunes, ainsi que d'autres éléments. Ces informations peuvent éclairer le diagnostic et la conception de programmes (voir les parties suivantes). Voir l'encadré 2.1. pour des exemples d'évaluation des jeunes et d'études de marché.

ENCADRÉ 2.1 Évaluations sur un échantillon de jeunes et études de marché

Évaluation concernant les jeunes et études de marché approfondies :

Peeters et al. 2009. *Youth Employment in Sierra Leone*. Washington, D.C., La Banque mondiale, http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/02/27/000334955_20090227091204/Rendered/PDF/476090PUB0Sier101Official0Use0Only1.pdf

Évaluations rapides concernant les jeunes :

Fondation internationale pour la jeunesse, 2010, *Building on Hope: Findings from a Rapid Community Appraisal in Jordan*, Baltimore, Fondation internationale pour la jeunesse, http://www.iyfnet.org/document/1059

Fondation internationale pour la jeunesse, 2011. *YouthMap Senegal – Youth Assessment: The Road Ahead*, Baltimore, Fondation internationale pour la jeunesse, http://www.iyfnet.org/document/1820

Diagnostic : quels sont les éléments déterminants dans l'obtention de résultats pour les jeunes ?

Que fait-on après avoir évalué les jeunes et mené une étude de marché ? Imaginons que nous avons conclu que les jeunes dans le pays, la région ou la ville où nous travaillons sont touchés, de manière disproportionnée, par le chômage et le sous-emploi. La plupart des jeunes âgés de 16 à 24 ans ne vont pas à l'école et n'ont pas de travail. Les filles semblent particulièrement touchées. Que devons-nous donc faire ? Quelle intervention proposer ?

En fait, il est trop tôt pour poser ces questions. Avant d'envisager une intervention, nous devons savoir *pourquoi* ces jeunes ne peuvent pas trouver de travail ou lancer leur propre affaire. Qu'est-ce qui les empêche d'entrer sur le marché du travail et de gagner leur vie pour eux et pour leur famille ? À quels obstacles se heurtent-ils ?

Imaginons que nous mettions en place des cours de formation technique pour les jeunes filles. Cette formation pourrait servir à améliorer les compétences artisanales et informatiques sur la

base d'un modèle concluant que notre organisation a mis en œuvre ailleurs. Et si ces compétences techniques n'étaient pas le problème ? Et si le réel problème pour ces jeunes filles était de ne pas savoir comment chercher un emploi et postuler, problème qui s'ajoute aux obstacles sociaux qui découragent les filles de travailler en dehors de chez elles ? Si nous ne disposons pas de suffisamment d'éléments ou que nous posons de mauvaises hypothèses sur les raisons pour lesquelles les jeunes ne peuvent pas trouver de travail, notre projet ne s'attaquera probablement pas aux causes du problème et échouera. Dans ce cas, le suivi et l'évaluation ne feront que confirmer une évidence.

Lors de la conception d'un programme d'emploi et de développement des entreprises pour les jeunes en Indonésie, une ONG locale n'a pas entièrement diagnostiqué les problèmes psychosociaux sous-jacents et les obstacles économiques auxquels les jeunes étaient confrontés depuis le tremblement de terre dévastateur dans la région. Il avait non seulement fait de nombreuses victimes et déplacé plus de 50 000 personnes mais aussi détruit les moyens de subsistance. Les jeunes, après ce traumatisme, n'ont pas eu accès à des services de conseils essentiels et le besoin d'un soutien psychosocial continu n'a même pas été reconnu. Même si les visites de suivi ont montré qu'il y avait eu des créations d'emplois, le programme n'a pas atteint tous ses objectifs parce que les jeunes ont continué de souffrir de dépression et qu'ils ont eu du mal à adopter les nouvelles technologies ou à prendre des mesures qui auraient pu augmenter la rentabilité de leurs petites entreprises.

Comprendre les obstacles à long et court terme

Il est essentiel de comprendre les obstacles sous-jacents qui peuvent limiter l'accès des jeunes au marché du travail et aux activités créatrices de revenus. Nous conceptualisons ici ces obstacles en deux groupes : ceux qui ne peuvent pas être levés pendant la durée de la plupart des programmes (obstacles à long terme) et ceux qui peuvent l'être (obstacles à court terme).

Obstacles à long terme. Les problèmes institutionnels et macro-économiques ne peuvent être rapidement résolus et vont fort peu probablement être influencés par des projets locaux individuels, dont la durée d'exécution est généralement de trois à cinq ans. Toutefois, il est important d'examiner les obstacles à long terme car ils représentent le contexte élargi de notre intervention. L'encadré 2.2. présente le cadre MILES, un aperçu des éléments déterminants structurels de la création d'emplois.

ENCADRÉ 2.2 Le cadre MILES

Stabilité macro-économique et politique. Les entrepreneurs ont besoin d'un cadre macro-économique solide dans lequel faire évoluer leur entreprise et créer de nouveaux emplois.

Climat propice aux investissements, institutions et infrastructures. Les entreprises s'étendront et créeront des emplois dans le secteur formel lorsque les coûts de l'entreprise (selon les réglementations en vigueur, le poids de l'impôt et la faiblesse des infrastructures) seront faibles et prévisibles.

Réglementation du marché du travail et institutions. Des réglementations rigoureuses sont essentielles pour l'employeur et le travailleur afin que leurs relations de travail soient productives et engagées sur le long terme.

Éducation et compétences. Les emplois hautement productifs sont invariablement liés à une bonne éducation formelle et des compétences adéquates sont nécessaires pour toutes les tranches d'âge.

Protection sociale. Un programme de protection sociale solide et équilibré protège le revenu des travailleurs des chocs sur l'emploi.

Source : Banque mondiale (2007c, p. 8–10).

Obstacles à court terme. Étant donné qu'il est difficile de changer la plupart des obstacles structurels à la création d'emploi et de sources de revenus, il est généralement plus réaliste pour les agents du développement de se concentrer sur les obstacles que l'on peut faire tomber en peu de temps. Parmi eux, cinq grandes catégories ressortent (Cunningham, Sanchez-Puerta et Wuermli 2010) :

- **Obstacles du côté des prestataires :** les jeunes n'ont pas les compétences professionnelles nécessaires sur le marché local. Ils ne savent parfois ni lire ni compter, ou ils n'ont pas les compétences techniques, comportementales ou entrepreneuriales nécessaires. Ils peuvent également se heurter à des obstacles non liés à leurs compétences, par exemple des problèmes psychosociaux, qui peuvent affecter leur employabilité (Rossiasco et al. 2010).

- **Obstacles du côté de la demande :** les employeurs recherchent peu une main-d'œuvre jeune du fait de la situation macroéconomique, par exemple le ralentissement de la croissance de l'emploi, et microéconomique, telle la discrimination de la part de l'employeur.
- **Obstacles en matière de création d'entreprises :** l'entrepreneuriat des jeunes se heurte à un accès insuffisant aux capitaux financiers, aux terrains ou aux réseaux sociaux.
- **Obstacles liés à l'intermédiation sur le marché du travail :** souvent, les jeunes ne disposent pas d'informations pertinentes et précises sur les vacances de postes et sur les qualifications demandées, ou ils ne savent pas présenter correctement leurs compétences à des employeurs potentiels.
- **Obstacles sociaux :** les normes sociales ou coutumes peuvent limiter le développement des compétences ou l'entrée sur le marché du travail de certains groupes, par exemple les filles, les jeunes autochtones, etc.

Le tableau 2.1 présente les obstacles possibles. Dans les faits, la difficulté est de définir les plus pertinents dans notre contexte local et de les classer par ordre d'importance. Chaque sous-groupe de population sera confronté à un groupe d'obstacles différent. Par exemple, les jeunes femmes dans le Rwanda rural vivent dans une économie à faible croissance, n'auront pas suffisamment de compétences, seront confrontées à une profonde discrimination de la part des employeurs et se heurteront aux normes imposées aux femmes, tandis que les hommes à faible revenu dans le Chili urbain se heurteront plutôt au manque d'information sur les possibilités d'emploi, aux difficultés de présenter leurs compétences à d'éventuels employeurs et à une inadéquation des compétences techniques ou personnelles (Cunningham, Sanchez-Puerta et Wuermli 2010). La liste d'obstacles auxquels notre population cible se heurte doit être définie par des analyses portant sur la jeunesse et des études de marché, comme décrit précédemment.

TABLEAU 2.1 Obstacles à court terme pour les jeunes sur le marché du travail

Obstacles		Description	Sources d'information
Du côté des prestataires : compétences pertinentes pour l'emploi et autres obstacles du côté des prestataires	Compétences de base insuffisantes	• L'écriture et le calcul sont le socle de la communication et des autres processus de développement des compétences. • Les jeunes en situation d'après-conflit peuvent être particulièrement touchés.	• Évaluation des compétences, notamment certification de compétences (par exemple PISA, TIMSS). • Évaluation du système éducatif (SABER). • Statistiques sur la réussite éducative. • Programmes scolaires (qu'enseigne-t-on ?). • Enquêtes existantes sur les employeurs. • Études de marché/de secteur.
	Compétences techniques insuffisantes ou inadéquates	• Les compétences propres au commerce ou à l'emploi vont des compétences manuelles à la maîtrise de l'outil informatique.	
	Compétences comportementales insuffisantes	• Les compétences comportementales - compétences générales - sont un éventail de qualités telles que la motivation, la résolution de problèmes, la communication, la gestion du temps et la capacité à travailler avec d'autres personnes. • Les compétences comportementales sont de plus en plus valorisées par les employeurs du monde entier.	
	Compétences entrepreneuriales insuffisantes	• La créativité, qui permet d'inventer ou d'adopter un nouveau produit ou processus, et les compétences commerciales pour commercialiser cette idée sont essentielles pour les employés et le travailleur indépendant.	
	Obstacles non liés aux compétences	• Autres obstacles qui peuvent empêcher un jeune d'accepter un travail ou de chercher du travail, par exemple le coût du transport, un enfant à charge, etc. • La santé d'un jeune, en particulier sa santé mentale, peut avoir une influence sur son employabilité. Par exemple, la dépression ou l'anxiété peut toucher ses compétences comportementales.	• Évaluation des jeunes • Évaluations ou études spécialisées sur la santé et la santé mentale

TABLEAU 2.1 (SUITE) Obstacles à court terme pour les jeunes sur le marché du travail

Obstacles		Description	Sources d'information
Du côté des demandeurs : demande de main-d'œuvre insuffisante	Lente croissance d'emplois	• Trop peu de nouveaux emplois sont créés (petit secteur formel). • Une croissance lente est souvent due à des facteurs économiques d'ensemble tels qu'un environnement difficile en matière d'investissement ou des chocs externes, par exemple des catastrophes naturelles, une guerre ou une modification soudaine de l'économie mondiale.	• Données relatives au PIB (par secteur). • Statistiques du marché du travail (notamment la part de l'économie formelle/informelle). • Enquêtes existantes sur les employeurs. • Étude de marché.
	Discrimination de la part des employeurs	• Les employeurs peuvent avoir des préjugés à l'égard des jeunes et croire, par exemple, que les jeunes sont moins fiables, moins dignes de confiance ou moins compétents que les personnes plus âgées. • Outre ces préjugés, les préférences en matière de recrutement peuvent être relatives au sexe, à l'appartenance raciale, ethnique ou religieuse.	• Enquêtes existantes sur les employeurs. • Étude de marché. • Études sociologiques.
Création d'entreprises : obstacles propres à la création d'entreprises	Accès aux capitaux financier, naturel et social insuffisant	• Des limites à l'auto-emploi peuvent comprendre l'absence de compétences entrepreneuriales et l'accès insuffisant à des fonds, terrains ou réseaux commerciaux.	• Étude de marché. • Statistiques bancaires/relatives à la micro-finance (produits d'épargne, exigences en matière de garantie, etc.). • Droits de propriété.
Intermédiation : obstacles liés à la recherche d'emploi	Adéquation professionnelle	• Souvent, les jeunes n'ont pas les réseaux qui permettent de prendre connaissance des emplois disponibles.	• Évaluation des jeunes. • Étude de marché.
	Signalement des compétences	• Les jeunes peuvent avoir les bonnes compétences mais peuvent avoir du mal à les présenter aux employeurs potentiels (par exemple par la preuve d'une expérience préalable ou par des certificats).	• Évaluation des jeunes (diplôme, certificats, etc.). • Étude de marché.
Obstacles sociaux	Obstacles rencontrés par les groupes marginalisés (appartenance ethnique, genre, etc.)	• Les coutumes locales et les normes sociales peuvent dissuader certains groupes de personnes de rejoindre le marché du travail. • La ségrégation professionnelle peut répondre à des limites raciales, ethniques ou religieuses.	• Étude de marché et/ou enquêtes existantes sur les employeurs (préférences en matière de recrutement). • Rapports sur les droits de l'homme. • Études anthropologiques.

Source : d'après Cunningham, Sanchez-Puerto, and Wuermli (2010).

Objectifs et élaboration : Quel est notre but ? Comment l'atteindre ?

Étant donné les problèmes spécifiés, le groupe cible et les obstacles spécifiques auxquels se heurte toute amélioration des possibilités économiques, nous pouvons formuler les objectifs du programme et choisir parmi les interventions possibles. Définir clairement nos buts nous aidera à réfléchir aux résultats finaux de notre programme, à communiquer avec les donateurs et parties prenantes, à administrer l'intervention, ainsi qu'à suivre et évaluer nos travaux.

Fixer l'objectif de développement du projet

Il faut commencer par définir notre objectif de développement du projet (PDO)[1], qui représente ce que nous voulons réaliser, le résultat escompté ou planifié de notre intervention. Plusieurs tâches peuvent nous aider à développer notre objectif principal, par exemple la détermination claire du groupe cible, l'ampleur des changements escomptés et la durée (voir figure 2.2.). Plus l'objectif sera concret, plus il sera facile de suivre les avancées réalisées pour l'atteindre.

FIGURE 2.2 Comment développer les objectifs du projet

Selon les problèmes et obstacles spécifiques identifiés :

1. Spécifier le **résultat** escompté : ce que nous pensons réaliser et non ce que nous voulons faire.
2. Spécifier la **population cible.**
3. Spécifier **l'orientation et l'ampleur** des changements escomptés.
4. Spécifier la **durée** au cours de laquelle les changements escomptés se produiront.
5. Faire en sorte que l'objectif soit **mesurable**.
6. Faire en sorte que l'objectif **puisse être attribué** à l'intervention.

Lorsque nous définissons notre objectif de développement du projet, nous faisons souvent l'erreur de mettre l'accent sur ce que nous allons faire et non sur ce que nous voulons réaliser (voir point 1 de la figure 2.2.). Si le but final de notre intervention est d'améliorer les conditions de vie des jeunes, cela doit se voir dans notre objectif de projet. Le moyen d'y parvenir, par exemple en fournissant une aide psychosociale, une formation, un capital d'amorçage ou d'autres services, constitue le moyen et non la fin en soi. L'encadré 2.3. examine trois exemples d'objectif de développement du projet.

1 Les organisations utilisent différents termes pour qualifier leurs principaux objectifs de projets, tel but du projet, but final, fin.

ENCADRÉ 2.3 Définir un objectif de développement du projet

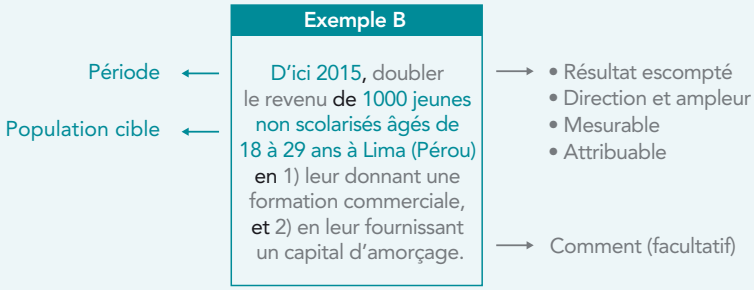
Exemple A	Exemple B	Exemple C
1000 jeunes Péruviens reçoivent une formation commerciale.	D'ici 2015, doubler le revenu de 1000 jeunes non scolarisés âgés de 18 à 29 ans à Lima (Pérou) en 1) leur donnant une formation commerciale, et 2) en leur fournissant un capital d'amorçage.	D'ici 2015, réduire le chômage des jeunes à Lima (Pérou) de 10 %.

Évaluation

Plusieurs détails nécessaires ne figurent pas dans cet objectif (période, groupe cible exact, etc.), ce qui le rend trop général. En outre, il ne mentionne aucun résultat escompté mais plutôt un service rendu. Former les jeunes n'entraînera pas nécessairement une amélioration de leur situation, par exemple de leurs connaissances, de leur statut d'emploi, ou de leur revenu. L'objectif du projet doit aller plus loin.

Cet exemple répond à tous les critères d'un bon objectif de projet. Il est concret et concerne une amélioration mesurable des conditions de vie du groupe cible.

Cet objectif, bien que spécifique et mesurable, ne peut être attribué à notre projet. Une seule intervention à la portée limitée ne pourra pas apporter les changements escomptés à haut niveau car le chômage des jeunes sera influencé par divers facteurs.


Exemple B
Période
Population cible
D'ici 2015, doubler le revenu de 1000 jeunes non scolarisés âgés de 18 à 29 ans à Lima (Pérou) en 1) leur donnant une formation commerciale, et 2) en leur fournissant un capital d'amorçage.
• Résultat escompté
• Direction et ampleur
• Mesurable
• Attribuable
Comment (facultatif)

Fixer des objectifs institutionnels

Outre la définition de l'objectif de développement du projet, nous pouvons également souhaiter définir des objectifs institutionnels. Bien qu'ils soient liés à notre intervention, ils peuvent ne pas concerner directement notre principal groupe cible. Par exemple, les objectifs institutionnels peuvent être internes à notre organisation et concerner par exemple les enseignements tirés du projet en matière d'élaboration et de mise en œuvre. Les objectifs institutionnels peuvent également concerner l'environnement du projet, par exemple la constitution de partenariats, le renforcement de la volonté politique ou l'amélioration de la participation des parties prenantes. Tous ces points sont importants et doivent être définis dès le début du projet.

Définir l'intervention

Avec un objectif clair en tête, nous pouvons définir la portée de l'intervention qui nous permettra de réaliser les objectifs fixés. Naturellement, le choix du programme devrait directement découler de la définition des obstacles définis à l'étape précédente, c'est-à-dire que nous devons choisir une intervention s'attaquant explicitement aux causes sous-jacentes qui empêchent les jeunes de développer les compétences nécessaires et de pouvoir gagner décemment leur vie pour eux et leur famille.

Élaboration de programmes fondés sur des données

Lorsque l'on élabore une intervention, il faut se pencher sur les connaissances actuelles relatives aux différents programmes possibles. Par exemple, pour faire tomber les obstacles liés à la création d'entreprises par les jeunes, on peut souhaiter mettre en œuvre un programme de promotion des entreprises créées par des jeunes. Mais à quoi l'intervention devrait exactement ressembler ? Imaginons que nous avons pu confirmer que les obstacles financiers constituaient les obstacles majeurs à la création d'entreprises. Le programme doit-il allouer des subventions ou des prêts ? Doit-il cibler les jeunes les plus jeunes ou les plus âgés ? Les mieux formés ou les moins bien formés ? De plus, un soutien financier sera-t-il suffisant ou faut-il y associer d'autres services d'appui tels que la formation, le parrainage et l'aide à la création d'entreprise ?

Pour répondre à ces questions difficiles, les administrateurs de programmes doivent se pencher sur les bases de données existantes. Souvent, nous avons tendance (ou les organisations pour lesquels nous travaillons) à favoriser certains types de projets selon nos préférences ou expériences. Pourtant, pour élaborer des projets de très bonne qualité, il est important de se pencher sur les connaissances théoriques et empiriques existantes en matière d'élaboration de programmes relatifs aux moyens de subsistance des jeunes (la partie consacrée aux ressources dans le présent document inclut des références à des revues universitaires, des bases de données et des évaluations d'impact terminées et en cours). Si les données disponibles confirment la solution pour laquelle nous penchions, nous pouvons en défendre le bien-fondé. Si, en revanche, elles indiquent d'importantes limites à une intervention, nous économiserons du temps et de l'argent si nous intégrons les enseignements tirés à la nouvelle initiative.

Le tableau 2.2. donne des exemples d'intervention dont le bilan est bon d'après les évaluations d'impact précédentes ou les données de suivi. S'appuyer sur ces programmes aidera à mieux concevoir les interventions et à les rendre plus crédibles. Lorsqu'il y a peu ou pas de données, l'intervention proposée n'est pas pour autant vouée à l'échec. En réalité, les démarches novatrices ne pourront par définition s'appuyer sur aucun précédent. Toutefois, lorsque nous menons des interventions qui ne s'appuient pas sur une base de données solide, nous devons toujours être conscients de leur caractère expérimental et ne pas prendre des résultats positifs pour acquis. C'est là qu'une évaluation rigoureuse sera particulièrement importante.

[Conseil]

Le Youth Employment Inventory (Inventaire sur l'emploi des jeunes, www.youth-employment-inventory.org) est un guichet unique pour les interventions terminées et en cours auprès des jeunes. Cette base de données dynamique permet de parcourir les interventions et évaluations et de les sélectionner par catégorie, ce qui permet aux utilisateurs de chercher les données disponibles sur un type de projet particulier.

TABLEAU 2.2 Liste des interventions fondées sur des données, par obstacle

Obstacle	Intervention fondée sur des données probantes	Intervention fondée sur des données mélangées
Compétences de base insuffisantes	• Informations sur la valeur de l'éducation	• Programmes d'école de la deuxième chance
Compétences techniques inadéquates	• Formations complémentaires/programmes complets • Informations sur le retour aux spécialités techniques	• Formation en cours d'emploi
Compétences comportementales inadéquates	Sans objet	• Formation aux compétences comportementales/générales
Compétences entrepreneuriales insuffisantes	Sans objet	• Formation entrepreneuriale
Économie à lente création d'emploi	• Subventions salariales et aides à la formation	• Programmes de service public • Travaux publics à forte intensité de main d'œuvre
Discrimination de la part des employeurs	• Programmes de discrimination positive	• Aides aux employeurs qui embauchent des personnes appartenant aux groupes cibles. • Parrainage des employés.
Accès aux capitaux financiers, naturels et sociaux insuffisants	• Programmes complets sur l'entrepreneuriat	• Micro-finance
Placement	• Services d'emploi	• Échange d'information grâce aux technologies
Signalement des compétences	Sans objet	• Certification des compétences • Accréditation des centres de formation
Obstacles rencontrés par les groupes exclus (ethnie, sexe, etc.)	• Participation des groupes exclus cibles aux programmes • Formation aux compétences non traditionnelles • Espaces de formation/d'emploi sûrs pour des groupes spécifiques	• Contenu/conception du programme adapté aux besoins spécifiques des groupes exclus

Source : d'après Cunningham, Sanchez-Puerta, and Wuermli (2010).

[Conseil]

Si l'on considère une vue d'ensemble du développement des jeunes, les stratégies de promotion des moyens de subsistance doivent être considérées au sens large, à savoir ce dont les jeunes ont besoin pour entrer dans l'âge adulte. Par exemple, le Search Institute présente quarante atouts internes et externes qu'ont les jeunes et qui peuvent être renforcés pour encourager leur avancée de l'avant.

Pour plus d'informations, veuillez consulter la page : http://www.search-institute.org/developmental-assets

Étant donnée la diversité économique, sociale, institutionnelle et administrative dans un même pays et entre les pays, ainsi que les besoins propres au groupe cible, toutes les interventions du tableau 2.2. ne seront pas nécessairement réalisables dans un contexte donné. Évaluer si les conditions préalables existent dans le pays ou le marché du travail cible et, dans la négative, si la conception du programme peut être revue pour permettre de mener l'intervention (Cunningham, Sanchez-Puerta et Wuermli 2010).

Lien entre conception et évaluation du programme

Enfin, il est important de reconnaître l'existence de liens étroits entre conception et évaluation du programme. Comme nous l'avons vu dans la note 1, l'un des principaux rôles de l'évaluation est de soutenir l'apprentissage, puis la planification. L'utilité et la faisabilité de l'évaluation dépendent donc beaucoup de la qualité de la première ébauche du programme. Gardez toujours à l'esprit les points suivants :

- **Une évaluation ne peut compenser une mauvaise conception.** Une évaluation a posteriori ne remplace pas la réflexion a priori. Un programme bien élaboré ancré sur des recherches existantes et les expériences vécues est le meilleur moyen de faire réussir le programme.
- **La stratégie d'évaluation dépendra des lacunes identifiées dans les connaissances pendant la conception du programme.** Connaître la base de données et définir les éventuelles

lacunes en matière de connaissances sont des facteurs importants pour choisir la bonne stratégie d'évaluation par la suite. Par exemple, les évaluations d'impact seront particulièrement utiles pour des programmes novateurs et jamais testés qui permettent de combler des lacunes.

- **Une bonne conception de programme peut faciliter l'évaluation.** Certains programmes sont plus faciles à évaluer que d'autres. Par exemple, si une évaluation d'impact n'est pas prévue dès la conception du programme, les instruments disponibles pour mener l'évaluation peuvent être très limités (voir note 6). Choisir des critères de ciblage clairs, équitables et transparents, tels que la répartition aléatoire pour les programmes recevant un grand nombre de candidats ou les scores d'admissibilité, peuvent grandement faciliter l'évaluation. Par conséquent, même s'il y a plusieurs manières acceptables de mener un programme particulier, il peut être sage d'anticiper et de concevoir un programme selon un modèle adapté à l'évaluation.

Points essentiels

1. Puisque l'utilité du suivi et de l'évaluation ne dépend que de la qualité de la conception du projet, nous devons garantir des normes élevées de planification et conception de nos interventions.
2. Pour concevoir des projets de qualité, nous devons comprendre les jeunes et le contexte dans lequel ils vivent. Pour ce faire, des évaluations des jeunes et des études de marché sont nécessaires pour saisir la complexité des facteurs environnementaux qui influencent le bien-être des jeunes et les possibilités qui s'offrent à eux.
3. Il est essentiel de diagnostiquer les facteurs sous-jacents qui empêchent les jeunes d'accéder à l'emploi et au revenu. Sans savoir exactement ce qui limite leurs possibilités, il est impossible de concevoir une intervention s'attaquant aux obstacles pertinents.
4. Lorsqu'une intervention est conçue pour réaliser l'objectif de développement du projet énoncé, consulter les données théoriques et empiriques existantes augmente les chances de réussite et évite de coûteuses erreurs.

Étude de cas du NUSAF : Examen de la conception du projet

Analyse du problème

Les taux élevés de chômage et de sous-emploi des jeunes sont des problèmes persistants qui constituent une priorité pour de nombreux gouvernements de pays à faible revenu et à revenu intermédiaire. Cela est aussi vrai pour le Gouvernement ougandais, qui s'efforce de réduire les risques de conflit dans le nord du pays.

Diagnostic

En Afrique en général, et dans le nord de l'Ouganda en particulier, il n'existe pratiquement aucune possibilité d'emploi formel car il n'existe pas d'entreprises privées. L'absence de possibilités d'emploi, le faible niveau de compétences et les obstacles à la création d'emploi font que le Programme visant à améliorer les perspectives d'emploi des jeunes a décidé de mettre l'accent sur un programme complet consacré à l'entrepreneuriat qui fournirait une formation professionnelle, des aides en espèces et d'autres services d'appui.

(suite page suivante)

Étude de cas du NUSAF : Examen de la conception du projet (suite)

Objectifs et conception

Le Programme avait pour objectif de fournir, avant 2010, au moins 8 000 emplois à des jeunes de 15 à 35 ans dans le nord de l'Ouganda en promouvant des entreprises fondées sur les compétences et le renforcement des installations de formations (l'ampleur escomptée des effets sur l'emploi n'a pas été spécifiée).

Outre le principal objectif, le Programme a ciblé plusieurs objectifs secondaires, notamment l'amélioration des interactions sociales des jeunes dans leur communauté et la diminution de la détresse psychologique causée par le récent conflit. Plus largement, le programme visait à contribuer au bien-être général des jeunes et de leurs ménages, en améliorant la santé et la qualité de vie, en apportant une croissance économique durable et en réduisant ainsi le risque de conflits dans le nord de l'Ouganda.

Source : Blattman, Fiala, and Martinez (2011).

À lire

Bidwell, K., Galbraith, C., et al. 2008. *Market Assessment Toolkit for Vocational Training Providers and Youth*. New York: Women's Commission for Refugee Women and Children and Columbia University School of International and Public Affairs. http://www.womensrefugeecommission.org/docs/ug_ysl_toolkit.pdf

Cunningham, W., Sanchez-Puerta, M. L., et Wuermli, A. 2010. *Active Labor Market Policies for Youth: A Framework to Guide Youth Employment Interventions.* Washington, DC: The World Bank. http://siteresources.worldbank.org/INTLM/214578-1103128720951/22795057/EPPNoteNo16_Eng.pdf

Fonds d'équipement des Nations Unies, 2011, *Listening to Youth: Market Research to Design Financial and Non-Financial Services for Youth in Sub-Saharan Africa.* New York: UNCDF. http://www.uncdf.org/english/microfinance/uploads/other/Listening%20to%20Youth-YouthStart%20Market%20Research.pdf

Women's Refugee Commission. 2009. *Building Livelihoods: A Field Manual for Practitioners in Humanitarian Settings*. New York: WRC. http://www.womensrefugeecommission.org/docs/livelihoods_manual.pdf

Notes

NOTE 3 : Mettre sur pied un système de suivi

Ce qui est mesuré est réalisé.

— Tom Peters

Une bonne évaluation ne peut être menée qu'avec un bon système de suivi. En outre, concevoir un bon système de suivi améliorera probablement la qualité générale de la conception de notre projet et en facilitera la gestion. La présente note résume les principales étapes de la construction du système de suivi qui devrait accompagner tout projet, qu'une évaluation soit menée ou non. Comme nous allons le voir, chaque projet doit être au moins doté des instruments de suivi suivants :

- Une chaîne de résultats
- Un cadre logique
- Un processus de collecte et d'analyse d'informations et d'application des conclusions

Pourquoi avons-nous besoin d'un système de suivi ?

Le suivi apporte des informations internes et externes, de manière continue, aux administrateurs de programmes, qui peuvent ainsi faire le point sur les avancées prévues et réelles. Lorsqu'une irrégularité ou un manque d'efficacité est repéré, on peut redresser la barre en temps voulu. Les activités de suivi consistent à collecter et à analyser les données pour vérifier que les ressources sont utilisées comme prévu, que les activités sont mises en œuvre comme prévu, que les produits et services prévus sont fournis et que les bénéficiaires visés sont atteints (Savedoff, Levine et Birdsall 2006). Un suivi efficace doit être associé à tout projet, qu'il fasse l'objet d'une évaluation ou non.

En outre, le suivi assoit l'évaluation d'une intervention. En pratique, il est difficile d'effectuer une évaluation de qualité sans informations correctes sur la mise en œuvre réelle. En l'absence d'informations fiables sur l'avancée et la qualité de la mise en œuvre, n'importe quelle évaluation risque de mal interpréter les raisons de la réussite ou de l'échec d'un projet.

Il est difficile, en assurant le suivi d'une intervention, de :

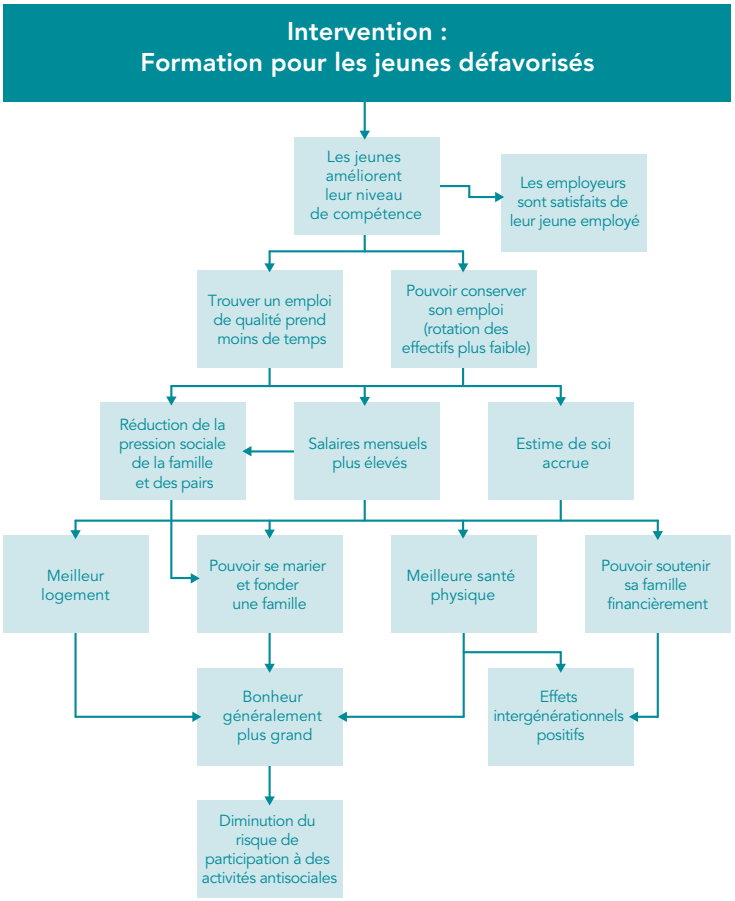
- Définir la *logique de l'intervention*, qui comprend la fixation d'objectifs au-delà de l'objectif de développement du projet, à tous les niveaux de mise en œuvre et de résultats.
- Déterminer les *indicateurs clés, les mécanismes de collecte de données, et les hypothèses* qui peuvent servir à mesurer les avancées réalisées par rapport aux objectifs fixés.
- Créer un *système de suivi et de rapport* pour suivre les avancées sur la voie de la réalisation des objectifs fixés et transmettre des informations aux administrateurs de programmes et autres parties prenantes.

Définir la logique de l'intervention

Lien entre conception du projet et projet théorique

Toute conception de projet comprend une théorie de changement. Comme nous l'avons vu dans la note 2, on s'attend généralement à ce qu'un projet aide à améliorer les conditions de vie de notre groupe cible en s'attaquant à un groupe particulier d'obstacles et de contraintes auxquels se heurtent ces jeunes. Cela étant, nous disposons d'un ensemble d'hypothèses sur la façon dont les activités du projet apporteront un changement positif et sur les raisons de cette amélioration. Pourquoi pensons-nous que former les jeunes améliorera la situation sur le marché du travail ? Pourquoi pensons-nous qu'aider les entreprises de jeunes réduira la pauvreté ? Pour confirmer la pertinence de notre intervention, la théorie qui la sous-tend doit être claire (voir figure 3.1.).

FIGURE 3.1 Théorie sous-tendant l'intervention d'un projet sur les moyens de subsistance des jeunes


Intervention :
Formation pour les jeunes défavorisés
Les jeunes améliorent leur niveau de compétence
Les employeurs sont satisfaits de leur jeune employé
Trouver un emploi de qualité prend moins de temps
Pouvoir conserver son emploi (rotation des effectifs plus faible)
Réduction de la pression sociale de la famille et des pairs
Salaires mensuels plus élevés
Estime de soi accrue
Meilleur logement
Pouvoir se marier et fonder une famille
Meilleure santé physique
Pouvoir soutenir sa famille financièrement
Bonheur généralement plus grand
Effets intergénérationnels positifs
Diminution du risque de participation à des activités antisociales

Les agents d'exécution devraient formuler une théorie du changement pour chaque intervention. Dans l'idéal, elle est formulée au début de la phase de conception du projet, lorsque toutes les parties prenantes peuvent être réunies pour convenir d'une vision commune du projet, de ses objectifs concrets et des mesures à prendre pour y parvenir (Gertler et al. 2011). Selon Taylor-Powell (2005), l'utilisation d'une théorie du changement aide l'administrateur du projet et l'évaluateur :

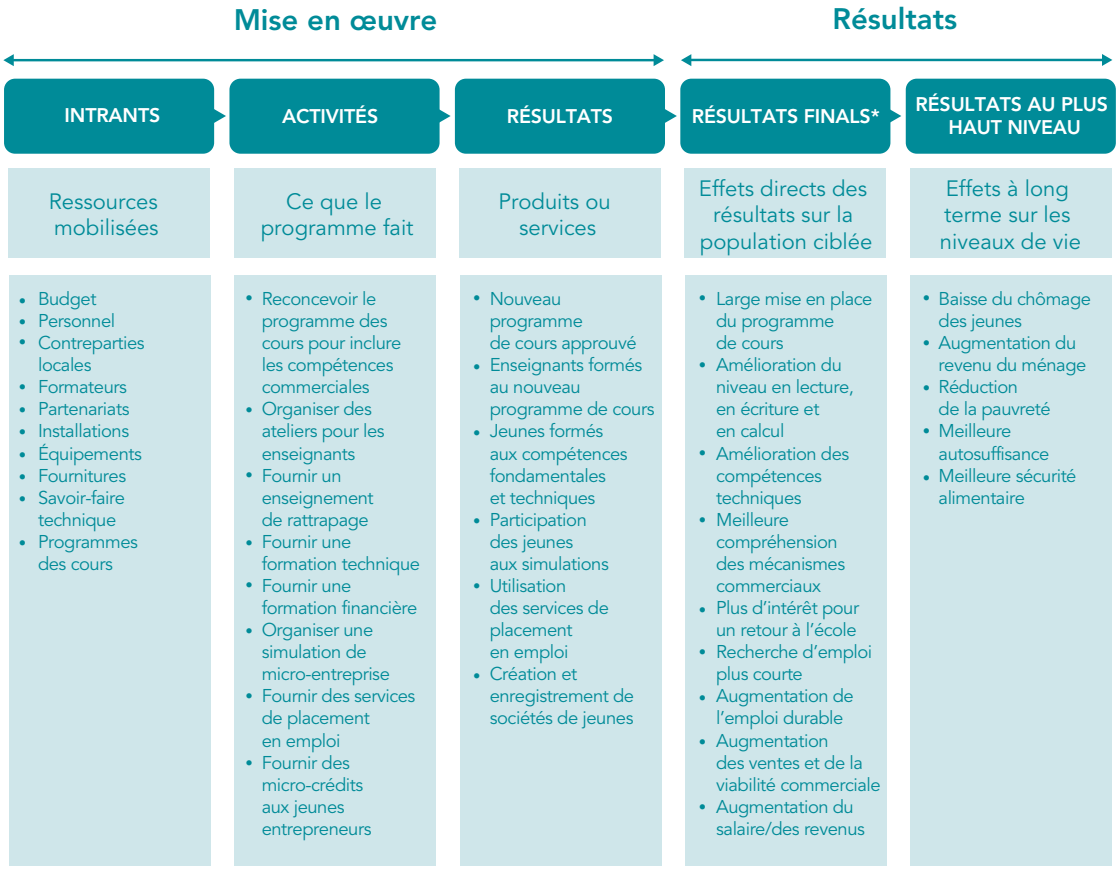
- En faisant mieux comprendre le projet et en offrant une langue commune ;
- En aidant à faire la différence entre « ce que nous faisons » et « ce à quoi nous voulons parvenir » ;
- En améliorant la planification et la gestion ;
- En définissant les variables importantes à mesurer ;
- En apportant un socle aux évaluations approfondies.

Faire de la théorie une chaîne de résultats

En pratique, une théorie du changement peut être appliquée de plusieurs manières. Les applications courantes sont les modèles logiques, les cadres logiques, les modèles de réalisation ou les chaînes de résultats. L'idée est toujours la même : fournir aux parties prenantes une présentation « logique et plausible » de la façon dont l'intervention prévue peut entraîner les résultats escomptés (Gertler et al. 2011, p. 24 ; voir figure 3.2.). Par conséquent, ces applications présentent une séquence

d'événements qui font le lien entre les éléments dépendant directement du projet (ressources employées, activités mises en œuvre et produits et services fournis), les résultats finals escomptés et les objectifs du programme au plus haut niveau.

FIGURE 3.2 Éléments d'une chaîne de résultats et exemples


Mise en œuvre
Résultats
INTRANTS
ACTIVITÉS
RÉSULTATS
RÉSULTATS FINALS*
RÉSULTATS AU PLUS HAUT NIVEAU
Ressources mobilisées
Ce que le programme fait
Produits ou services
Effets directs des résultats sur la population ciblée
Effets à long terme sur les niveaux de vie
• Budget
• Personnel
• Contreparties locales
• Formateurs
• Partenariats
• Installations
• Équipements
• Fournitures
• Savoir-faire technique
• Programmes des cours
• Reconcevoir le programme des cours pour inclure les compétences commerciales
• Organiser des ateliers pour les enseignants
• Fournir un enseignement de rattrapage
• Fournir une formation technique
• Fournir une formation financière
• Organiser une simulation de micro-entreprise
• Fournir des services de placement en emploi
• Fournir des micro-crédits aux jeunes entrepreneurs
• Nouveau programme de cours approuvé
• Enseignants formés au nouveau programme de cours
• Jeunes formés aux compétences fondamentales et techniques
• Participation des jeunes aux simulations
• Utilisation des services de placement en emploi
• Création et enregistrement de sociétés de jeunes
• Large mise en place du programme de cours
• Amélioration du niveau en lecture, en écriture et en calcul
• Amélioration des compétences techniques
• Meilleure compréhension des mécanismes commerciaux
• Plus d'intérêt pour un retour à l'école
• Recherche d'emploi plus courte
• Augmentation de l'emploi durable
• Augmentation des ventes et de la viabilité commerciale
• Augmentation du salaire/des revenus
• Baisse du chômage des jeunes
• Augmentation du revenu du ménage
• Réduction de la pauvreté
• Meilleure autosuffisance
• Meilleure sécurité alimentaire

** Niveau de l'objectif de développement du projet*

[Définition]
Une **chaîne de résultats** est une séquence de ressources, d'activités et de services fournis qui doivent avoir une influence directe et sur le long terme sur notre population cible.

Notre processus planifié de mise en œuvre inclut les catégories suivantes, dont l'administrateur du programme est directement responsable :

- **Intrants**—les ressources dont dispose le projet, notamment le budget, le personnel, les partenaires et les équipements.
- **Activités**—les actions, processus, techniques, outils, événements et technologies du programme. Décrire ces activités avec un verbe d'action (*fournir, faciliter, réaliser, organiser,* etc.).
- **Résultats**—les produits et services fournis directement sous le contrôle de l'organisme d'exécution. Elles indiquent si un programme a été réalisé comme prévu. Les Résultats sont généralement exprimés comme étant des actions terminées (*formé, participé, utilisé, financé,* etc.).

Les résultats escomptés décrivent tous les effets que le programme devrait avoir dans les catégories suivantes :

- **Résultats finals**—effets à court et moyen terme (généralement pendant plusieurs mois et pour un maximum de deux ans) sur la population bénéficiaire grâce aux résultats du projet. Il peut

s'agir de changements dans les attitudes, les connaissances et les compétences, qui peuvent souvent avoir des effets relativement immédiats, ainsi que de changements dans les comportements, les statuts et tout ce qui s'y rapporte, ce qui peut prendre plus de temps. Les principaux résultats finals visés devraient être ceux définis dans l'objectif de développement du projet. Ils sont généralement exprimés au niveau individuel et indiquent un changement observable (*amélioré, accru, réduit,* etc.).

- **Résultats au plus haut niveau**—objectifs du projet à long terme, généralement liés aux niveaux de vie d'un point de vue général. Ils peuvent subir l'influence de plusieurs facteurs et ne sont généralement pas complètement sous le contrôle du programme. Ce niveau de la chaîne de résultats est souvent qualifié d'« impacts ». Nous préférons l'expression « résultats au plus haut niveau » pour éviter toute confusion avec le terme « impact » dans le contexte de l'évaluation d'impact (voir note 5).

[Conseil]

Même si cela n'est pas absolument nécessaire, il est souvent judicieux d'inclure vos objectifs institutionnels et activités sous-jacentes dans la chaîne de résultats.

Construire une chaîne de résultats

Définir le niveau d'observation

Concernant tant l'exécution que les résultats, nous pouvons souhaiter nous pencher sur d'autres niveaux que les jeunes. Nous pouvons être intéressés par les résultats ou les résultats finals au niveau des ménages, du groupe ou des installations (écoles, centres de formation professionnelle) ou du village/de la communauté.

Étudier la diversité des résultats finals possibles

Les interventions axées sur les moyens de subsistance des jeunes peuvent toucher plusieurs résultats finals qui concernent directement les possibilités économiques et le marché du travail. Selon l'intervention, il peut s'avérer utile de cibler et de mesurer plusieurs résultats finals s'ils correspondent à la logique et aux objectifs du projet. Les catégories courantes de résultats finals comprennent les éléments suivants :

- **Développement psychosocial**—mesure l'esprit, les émotions et le niveau de maturité d'un jeune. Les résultats finals peuvent avoir trait à l'estime de soi, à l'identité, à la confiance, à l'isolement ou au bien-être psychosocial.
- **Compétences**—niveau de base en lecture, écriture et calcul ; compétences techniques dans un commerce particulier (artisanat, mécanique, comptabilité, services au consommateur) ; connaissances de base (communication, travail en équipe, pensée critique, autogestion) ; et compétences entrepreneuriales (créativités, compétences commerciales).
- **Emploi et marché du travail**—utilisation que les bénéficiaires font de leur temps (entre les heures d'école, emploi salarié, auto-emploi, chômage, travail occasionnel) ; caractéristiques de l'emploi (type d'employeur ou d'affaire, nombre d'heures travaillées, revenus) ; et caractéristiques commerciales (profits, nombre d'employés, survie de l'entreprise, taux de remboursement du prêt).
- **Utilisation des services financiers**—accès des bénéficiaires aux services financiers et comportements en matière de banque, d'épargne, de gestion de la dette, de budget et de santé financière d'un point de vue général.
- **Comportements à risque**—attitudes et comportements liés à la consommation d'alcool, de tabac et de drogues, santé génésique (par exemple rapports sexuels non protégés, VIH/sida), crime et violence.

- **Formation de la famille**—attitudes et comportements concernant l'âge du mariage, le nombre d'enfants désirés et le nombre d'enfants nés.
- **Citoyenneté**—préférences des jeunes et actions en matière de vote aux élections locales ou nationales, participation à la communauté (par exemple par l'appartenance à un club ou par le bénévolat) et rôles d'animation.
- **Investissements dans le capital humain**—changements de la situation scolaire (retourner à l'école ou souhaiter y retourner), montant dépensé dans l'éducation ou la santé (pour soi ou pour autrui) et contributions intergénérationnelles (par exemple, vaccination et suivi de la croissance de ses enfants).
- **Autres**—les résultats finals supplémentaires peuvent porter sur les modes de consommation et d'alimentation, la création d'actifs, la mobilité et la migration, ainsi que les ménages et les relations communautaires.

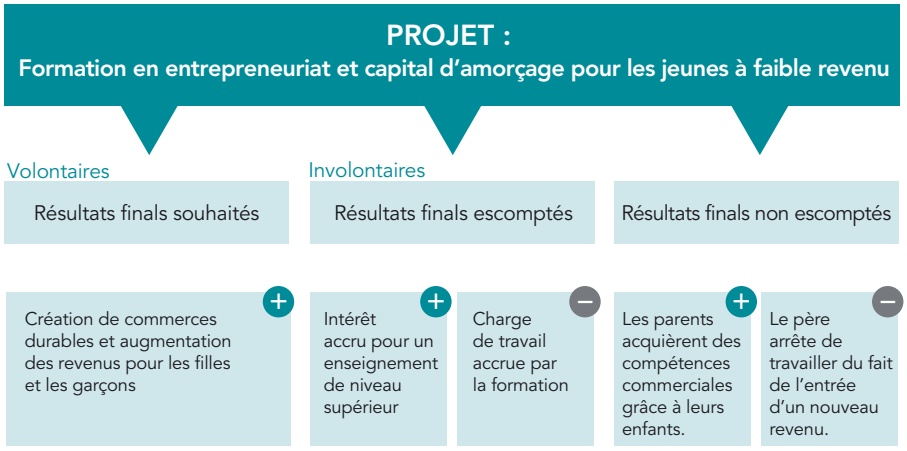
Au début des années 2000, le Conseil de population et Save the Children ont mis en œuvre le Programme Ishraq dans le nord rural de l'Égypte, créant ainsi des espaces adaptés aux filles pour dispenser des compétences fondamentales, construire des réseaux sociaux et promouvoir l'exercice de responsabilités et la confiance en soi. Il s'est avéré que les retombées du programme ont dépassé les adolescentes non scolarisées et atteint les parents des participantes. Les filles ont transmis à leur mère les informations apprises en cours, notamment en matière de santé. En outre, voir sa fille participer à la vie publique a un effet important sur la vision qu'a la mère de sa place dans la sphère publique. Grâce à la participation de leurs filles à Ishraq, les mères ont réalisé qu'elles aussi avaient le droit d'avoir accès aux services publics.

Source : Brady, Salem et Zibani (2007).

Tenir compte des résultats finals involontaires

Notre objectif de projet reflète le principal résultat final escompté de l'intervention. Toutefois, les projets de développement sont complexes et notre intervention peut avoir des effets *involontaires*. Certains d'entre eux peuvent être escomptés tandis que d'autres ne le sont pas et s'avèrent surprenants. Dans les deux cas, ces résultats finals peuvent être positifs ou négatifs (voir figure 3.3.). Il est important d'inclure ces résultats finals potentiels (voir les grandes catégories ci-dessus) dans la chaîne de résultats et dès lors de les identifier pour saisir, de manière réaliste, la logique complète de l'intervention et fournir le socle du suivi de tout mécanisme.

FIGURE 3.3 Résultats finals volontaires *vs* résultats finals involontaires


PROJET :
Formation en entrepreneuriat et capital d'amorçage pour les jeunes à faible revenu
Volontaires
Involontaires
Résultats finals souhaités
Résultats finals escomptés
Résultats finals non escomptés
Création de commerces durables et augmentation des revenus pour les filles et les garçons
Intérêt accru pour un enseignement de niveau supérieur
Charge de travail accrue par la formation
Les parents acquièrent des compétences commerciales grâce à leurs enfants.
Le père arrête de travailler du fait de l'entrée d'un nouveau revenu.

Source : d'après Hempel (2006).

Notre intervention peut par exemple avoir des effets secondaires parce que les participants transmettent leurs connaissances à leur famille ou aux membres de leur communauté qui en bénéficient à leur tour, indirectement. Nous souhaitons de toute évidence obtenir cet effet. Par ailleurs, il peut y avoir des effets préjudiciables inattendus : dans un projet sur l'entrepreneuriat, par exemple, certains jeunes sont endettés car leur affaire n'a pas survécu. Dans d'autres cas, lorsque les revenus des jeunes sont supérieurs grâce à notre intervention, les membres de la famille peuvent cesser de travailler ou se servir du revenu supplémentaire pour adopter davantage de comportements nocifs pour

la santé, par exemple la consommation d'alcool et de tabac. Dans ce cas, nous souhaitons également savoir si ces effets risquent de survenir. Faire des recherches sur des projets similaires aide souvent à déterminer l'éventail de résultats finals positifs et négatifs potentiels.

Éviter toute redondance dans les activités ou les résultats finals

Lors de l'élaboration de notre chaîne de résultats, il se peut que nous déterminions des activités ayant peu à voir avec notre principal objectif de projet. Pour que le projet soit bien défini ou efficace, les activités et résultats finals non essentiels doivent être abandonnés.

[Conseil]

Si le suivi des résultats finals involontaires risque de surcharger le cadre de résultats, les équipes du projet peuvent choisir de mettre l'accent sur le suivi des résultats finals involontaires et de se servir des évaluations pour en mesurer la portée.

Définir les indicateurs clés, les instruments de collecte des données et les hypothèses

Une fois la chaîne de résultats établie, comment savons-nous si ce qui a été prévu est réalisé ? L'une des plus grandes difficultés de l'élaboration d'un système de suivi est le choix du type d'information qui reflète le mieux si nous parvenons à réaliser nos objectifs. Nous essayons désormais de déterminer les indicateurs adéquats, les instruments de collecte des données et les hypothèses, des Intrants aux Résultats au plus haut niveau. Un cadre logique constitue une grille utile pour bien rendre compte de tous ces éléments (voir tableau 3.1.).

Étape 1 : Déterminer les indicateurs

Les indicateurs répondent à la question « Comment vais-je savoir ? ». Les indicateurs sont :

- Les aspects fondamentaux (ou descriptifs) de l'élément que nous souhaitons mesurer, même s'ils ne sont pas totalement représentatifs.
- Les signes tangibles qui montrent que quelque chose a été fait ou réalisé ; ce sont les moyens que nous qualifions de marques de notre réussite (Shapiro 2003).

Les indicateurs sont un élément essentiel d'un système de suivi car ils sont le moteur de la collecte de données, de leur analyse et de l'élaboration de rapports. Sans un ensemble clair d'indicateurs, les activités de suivi ou d'évaluation ne peuvent comparer les avancées réelles du programme à l'aune de ce qui a été prévu et convenu (Gosparini et al. 2004).

TABLEAU 3.1 Exemple de cadre logique pour un programme scolaire d'entrepreneuriat

		Étape 1	Étape 2			Étape 3
	Objectifs	**Indicateurs et cibles**	**Sources d'information**	**Fréquence**	**Partie responsable**	**Hypothèse**
Résultats au plus haut niveau	• Réduire le chômage des jeunes • Augmenter le revenu du ménage	• Taux de chômage local (%) • Revenu du ménage ($)	• Statistiques de l'emploi (niveau du ministère et de la ville) • Enquête sur les ménages	• Annuelle	• Équipe du programme	• Les nouvelles compétences sont demandées sur le marché du travail, qui les récompense.
Résultats finals	• Large mise en place du programme de cours • Meilleure compréhension des mécanismes commerciaux • Amélioration des compétences de base • Meilleure employabilité • Plus d'intérêt pour l'enseignement supérieur	Dans les six mois qui suivent la fin du programme : • 500 écoles utilisent le nouveau programme de cours. • Augmentation de moitié du nombre de réponses justes sur les connaissances commerciales après essai. • 70 % des élèves sont satisfaits du nouveau programme. • La vision qu'ont les enseignants et les parents des compétences de base augmente de 30 %. • La durée de la recherche d'emploi diminue de moitié et la satisfaction de l'employeur augmente de 30 %. • Augmentation de 5 % du nombre d'inscriptions à l'université.	• Entretien avec les autorités chargées de l'enseignement • Résultats des tests scolaires • Groupe de réflexion réunissant enseignants et parents • Étude sur les employeurs • Statistiques régionales sur l'inscription à l'école	• Biannuelle	• Équipe du programme (entretiens, collecte de données) • Consultant du programme (étude, groupe de réflexion)	• Le programme de cours est accepté par les autorités scolaires locales • Qualité de l'enseignement • Les jeunes peuvent aller régulièrement à l'école
Résultats	• Approbation du nouveau programme de cours • Formation des enseignants • Jeunes formés aux compétences commerciales	Avant la fin du programme : Le nouveau programme est approuvé par le ministère. • 500 enseignants ont été formés. • 10 000 jeunes ont terminé la formation.	• Données du programme	• Bimensuelle	• Équipe du programme	• Les enseignants souhaitent être formés. • Les jeunes peuvent suivre la formation.
Activités	…	…	…	…	…	…

Note : Pour des raisons pratiques, nous avons omis la catégorie des activités et des Intrants, qui devraient généralement figurer dans le cadre logique.

Choisir les indicateurs pour toute la chaîne de résultats

Même si nous mettons l'accent sur les résultats de l'intervention, il est important de suivre les indicateurs d'exécution afin de déterminer si le projet a touché les bénéficiaires visés et s'il a été mené comme prévu. Sans ces indicateurs tout au long de la chaîne de résultats, une évaluation n'établira que la réalisation des résultats finals prévus sans pouvoir faire le lien entre le niveau atteint et la qualité de l'exécution du programme. Le tableau 3.2. donne des exemples d'indicateurs tout au long de la chaîne de résultats.

TABLEAU 3.2 Exemples d'indicateurs

Catégorie	Exemple de cible	Exemple d'indicateurs
Entrée	Deux formateurs et installations avec un budget de 10 000 dollars E.-U.	• Deux formateurs formés, équipés et déployés • Le coût du programme en dollars E.-U. s'inscrit dans le budget escompté
Activité	Former les jeunes aux compétences de base (20 heures)	• Nombre d'heures de formation fournies • Nombre de jeunes participants, par âge, sexe et niveau d'éducation • Date de la formation
Résultats	100 jeunes ont participé à la formation	• Nombre de jeunes ayant fini la formation (par âge, sexe et niveau d'éducation)
Résultats finals	Davantage de connaissances en matière de communication efficace	D'ici la fin du programme : • Nombre et pourcentage de jeunes capables d'exprimer clairement des idées, selon une grille de résultats au test prédéterminée • Nombre et pourcentage de jeunes ayant amélioré leurs compétences en matière de communication verbale et non verbale, selon une grille de résultats au test prédéterminée • Nombre et pourcentage de jeunes qui disent se sentir à l'aise pour contacter un employeur.
Résultats au plus haut niveau	Augmentation du revenu des ménages	• D'ici 2015, le revenu moyen des ménages augmente de 20 % par rapport à la situation initiale.

Il faut être particulièrement vigilant pour définir de bons *indicateurs de réalisation*. Comme expliqué précédemment, les résultats finals issus des interventions dans le domaine des moyens de subsistance des jeunes peuvent être très variés et ne se limitent pas aux résultats finals sur le marché du travail. Nous devons donc choisir des indicateurs (développement psychosocial, compétences, emploi, etc.) parmi tous les domaines appropriés. Les domaines précis à mesurer dépendent bien sûr de l'objectif et de l'axe de l'intervention, ainsi que des objectifs d'apprentissage à atteindre. Par exemple, pour une intervention axée sur les compétences de base, il peut s'avérer utile de mesurer les compétences, les résultats finals sur le marché du travail et les comportements à risque. En revanche, un projet d'appui au placement en emploi peut être entièrement axé sur les résultats finals sur le marché du travail.

[Ressource en ligne]

Résultats finals choisis et indicateurs de résultats

http://www.iyfnet.org/gpye-m&e-resource1

Spécifier les indicateurs

Associer les autres parties prenantes. Si vous choisissez les indicateurs sans associer correctement les principales parties prenantes internes et externes, elles n'auront pas le sentiment d'être véritablement parties prenantes (Kusek et Rist 2004). Il faut collaborer avec les partenaires locaux et les acteurs communautaires pour parvenir à un ensemble de buts, objectifs et indicateurs de résultats pour le programme, fixés ensemble.

Choisir le bon nombre d'indicateurs. Puisque les indicateurs ne sont que des indications indirectes, il est courant d'en définir plusieurs pour chaque élément de la chaîne de résultats, en particulier concernant les résultats finals ou les résultats au plus haut niveau. Toutefois, un nombre trop important d'indicateurs compliquera inutilement notre système de suivi et pèsera sur la collecte et l'analyse de données, ainsi que sur l'établissement des rapports. Il est important de déterminer les indicateurs (entre un et trois) qui reflètent le mieux chaque élément de la chaîne de résultats.

Respecter les normes de qualité. Même s'il n'y a pas de principe absolu sur ce qui fait la qualité d'un indicateur, les caractéristiques SMART généralement citées peuvent être utiles (Gertler et al. 2011, p. 27). Un indicateur doit être :

- Spécifique, pour mesurer les informations requises le plus précisément possible,
- Mesurable, pour garantir que l'information soit rapidement obtenue,
- Attribuable, pour veiller à ce que chaque mesure soit liée à l'effort du projet,
- Réaliste, pour garantir que les données soient obtenues en temps voulu, à une fréquence et à un coût raisonnables ; et un indicateur doit
- Toucher la population cible de l'objectif.

Le choix que nous ferons en matière d'indicateurs sera en partie déterminé par nos capacités à collecter des données. Si un indicateur n'est pas mesurable ou qu'aucune information n'est disponible, il ne peut pas refléter l'avancée de nos objectifs. Si nous ne pouvons pas collectelir de données pour un indicateur choisi, nous devons le remplacer.

Établir une ligne de référence. La ligne de référence nous donne la valeur d'un indicateur au début ou, idéalement, juste avant la période de mise en œuvre. Connaître la valeur de référence de nos indicateurs nous permet de définir des cibles réalistes et de suivre les avancées futures par rapport à la situation initiale. Par exemple, si nous voulons suivre les revenus des participants au fil du temps, les données des formulaires d'enregistrement au programme pourront indiquer que le revenu moyen mensuel des participants au début du programme s'élevait à 100 dollars. Il s'agit de la valeur de référence qui peut servir de comparaison pour établir la hausse des revenus pendant et après notre intervention.

Définir les cibles. Si le calendrier, la quantité et la qualité ne figurent pas dans les indicateurs, on ne peut pas être entièrement sûr d'être sur la bonne voie et de parvenir à réaliser ses objectifs (Cooley 1989). Par exemple, si le résultat final escompté est une augmentation du revenu des ménages, l'indicateur peut être les revenus mensuels en dollars E.-U. Par conséquent, la cible peut être fixée à une augmentation de 30 % (quantité) de l'emploi dans le secteur formel (qualité) pendant trois ans (calendrier). Chaque indicateur ne devrait pas avoir plus d'une cible par période. Si fixer des cibles quantitatives précises est trop arbitraire, on peut les exprimer sous forme d'éventail.

Garantir la cohérence. Même si cela n'est pas toujours possible, pour garantir un suivi cohérent au fil du temps, nous devrions nous efforcer de conserver les indicateurs qui ont été convenus avant le lancement du projet. Cela étant, il n'est pas rare d'ajouter de nouveaux indicateurs et d'en abandonner d'autres à mesure que nous modifions la conception du programme et intégrons le système de suivi. Il est néanmoins essentiel de ne jamais oublier les objectifs originaux du projet. Le suivi et l'évaluation doivent être honnêtes. Si nous voyons que le projet ne réalisera pas son but original mais un autre (qui sera peut-être encore plus important), nous devons le reconnaître lors de l'établissement du rapport. Les indicateurs acceptés au début de l'intervention ne doivent être changés que si des critères objectifs justifient ce changement.

Le tableau 3.3. donne des exemples d'indicateurs pour des interventions axées sur les moyens de subsistance des jeunes à tous les niveaux de la chaîne de résultats. Il est parfois possible d'utiliser des indicateurs prédéfinis. Toutefois, il est important d'en examiner la pertinence pour le projet en question. Certains doivent être adaptés avant d'être inclus ou d'être ajoutés à d'autres, plus pertinents localement.

[Conseil]

Il est généralement judicieux de mener un essai sur les indicateurs pendant les premières phases d'une intervention avant de les intégrer au système de suivi. Cela mettra en avant leur niveau de performance et montrera s'ils recueillent toutes les informations que l'administrateur du projet et les autres parties prenantes souhaitent obtenir.

Résultat final à mesurer : *amélioration de l'employabilité des jeunes âgés de 18 à 24 ans*

Mauvais indicateur : Les jeunes trouveront plus facilement un emploi après l'intervention.

Bon indicateur : Nombre et pourcentage de jeunes âgés de 18 à 24 ans qui ont au moins deux propositions d'emploi, pour un salaire supérieur au salaire minimum dans leur domaine de formation, dans les trois mois qui suivent la fin du programme (contre aucune offre d'emploi auparavant).

TABLEAU 3.3 Exemples d'indicateurs pour les projets axés sur les moyens de subsistance des jeunes

Type de projet	Intrants	Activités	Résultats	Résultats finals	Résultats au plus haut niveau
Formation et développement de compétences	• Allocation budgétaire et dépenses (en dollars E.-U.) • Montant et part des fonds complémentaires levés • Nombre de membres de personnel du programme par niveau • Nombre de facilitateurs locaux sous contrat • Nombre d'organisations locales qui fournissent des contributions en nature	• Nombre d'ateliers offerts • Nombre d'heures de formation • Nombre de jeunes sélectionnés/ inscrits • Nombre d'employeurs offrant des stages • Nombre de stages disponibles	• Nombre et pourcentage de jeunes ayant suivi au moins 80 % de la formation • Nombre de certificats délivrés • Nombre de jeunes placés dans des stages • Durée moyenne des stages menés à terme (en semaines)	• Nombre et pourcentage de jeunes satisfaits du programme • Nombre et pourcentage de jeunes qui indiquent penser de manière plus critique et être plus à même de résoudre les problèmes • Nombre et pourcentage de jeunes recevant des offres d'emploi après un stage • Pourcentage d'employeurs locaux offrant des possibilités d'emplois aux jeunes	• Revenu du ménage (en dollars E.-U.) • Taux de chômage local des jeunes (%) • Niveau de consommation alimentaire par individu/ ménage (y compris les fruits et légumes) • Nombre et pourcentage de jeunes qui indiquent que leur maison/ appartement dispose des infrastructures de base (eau courante, électricité, etc.) • Nombre et pourcentage de jeunes qui indiquent une baisse des niveaux de conflit au cours de l'année écoulée
Emploi subventionné (par exemple, programmes de travaux publics et de services publics)	Idem	• Nombre de projets de mise au travail des bénéficiaires d'allocations, par type et lieu • Nombre de municipalités offrant des services publics/ menant des travaux publics	• Nombre de bénéficiaires employés dans chaque activité • Nombre d'emplois temporaires créés (par type et par secteur)	• Nombre et pourcentage de jeunes qui sont passés à un emploi formel en *n* mois • Heures et jours travaillés par semaine (par type d'activité) • Salaire moyen par heure/jour/mois	Idem

TABLEAU 3.3 (SUITE) Exemples d'indicateurs pour les projets axés sur les moyens de subsistance des jeunes

Type de projet	Intrants	Activités	Résultats	Résultats finals	Résultats au plus haut niveau
Services de l'emploi (par exemple aide au placement en emploi)	Idem	• Nombre de services de conseils d'orientation professionnelle créés (dans les bureaux, les écoles, etc.) • Nombre de sessions d'orientation professionnelle proposées • Nombre de forums étudiants et de forums pour l'emploi organisés	• Nombre de jeunes participant à des services de placement en emploi • Nombre et pourcentage de jeunes mis en rapport avec des employeurs • Nombre de sociétés et de jeunes participant aux forums locaux pour étudiants/pour l'emploi	• Nombre d'entretiens d'embauche par bénéficiaire • Nombre et pourcentage de jeunes employés *n* mois après l'intervention • Nombre et pourcentage de jeunes ayant conservé leur emploi pendant au moins *n* mois	Idem
Entreprises et entrepreneuriat de jeunes	Idem	• Nombre de compétitions de plans commerciaux organisées • Nombre d'heures de services d'appui fournies • Nombre moyen d'heures de conseil fournies par semaine/mois	• Nombre de jeunes présentant un plan commercial complet • Nombre d'entreprises de jeunes aidées par année • Nombre et pourcentage de jeunes parlant avec leur conseiller au moins une fois tous les 15 jours	• Nombre et pourcentage de jeunes qui ont lancé une nouvelle affaire • Nombre et pourcentage d'affaires enregistrées • Total des ventes la semaine dernière/le mois dernier • Nombre d'emplois créés • Pourcentage des bénéfices réinvestis	Idem
Services de financement incluant les jeunes	Idem	• Nombre d'ateliers organisés pour les établissements financiers participants • Lancement d'un programme de micro crédit pour les jeunes entrepreneurs • Création de comptes d'épargne pour les jeunes	• Nombre de personnel formé dans les établissements financiers partenaires • Nombre de prêts commerciaux accordés à des jeunes (par type d'entreprise) • Taille moyenne du prêt • Nombre de comptes d'épargne pour les jeunes ouverts	• Taux de remboursement annuel • Montant actuel de l'épargne 1) sur un compte bancaire, 2) dans un groupe d'épargne, 3) partout ailleurs • Nombre et pourcentage de jeunes qui épargnent dès qu'ils reçoivent de l'argent • Nombre et pourcentage de jeunes qui se disent plus satisfaits de leur situation financière	Idem

Étape 2 : Collecte de données

Le choix des indicateurs utilisés pour notre système de suivi dépend non seulement de la structure et des objectifs du projet mais aussi de la disponibilité des données, ainsi que du temps et des compétences nécessaires à leur collecte. Par données, on entend tout type d'information, et non uniquement les informations quantifiables.

Au Pakistan, l'organisation MEDA (Mennonite Economic Development Associates) suit ses projets de développement économique rural par un système de SMS. Les microentrepreneures et les propriétaires de petites entreprises envoient par SMS, chaque jour ou chaque semaine, le nombre de ventes réalisées.

Choisir une méthode de collecte des données

Il existe de grandes méthodes de collecte des données : quantitative et qualitative.

Les méthodes quantitatives visent à brosser un tableau objectivement mesurable de la situation, de manière strictement prédéfinie. Elles fournissent des informations sur la population visée, d'un point de vue fermé et quantitatif, et incluent des caractéristiques démographiques, socio-économiques ou d'autre nature. Elles sont généralement fondées sur des instruments structurés normalisés qui fournissent des données agrégées et facilitent l'analyse comparative. Les exemples courants sont les essais, les études et les recensements. Il faut des compétences statistiques pour mener des méthodes quantitatives.

Les méthodes qualitatives visent à faire comprendre comment pensent et se comportent les personnes, ainsi que les raisons de ces modes de pensée et de comportement. Les méthodes qualitatives s'efforcent de comprendre les événements du point de vue des acteurs, d'analyser comment différents groupes de personnes interprètent leurs expériences et construisent la réalité. Les exemples courants de méthodes qualitatives comprennent les entretiens non structurés ou semi-structurés, les groupes de discussion et l'observation directe des participants. Une formation en anthropologie ou en sociologie, ainsi qu'à l'administration d'instruments d'évaluation, est nécessaire pour mener des méthodes qualitatives. On peut généralement les mettre plus rapidement en œuvre que les méthodes quantitatives, et elles sont souvent moins chères.

Les règles qui régissent les statistiques sont transparentes et relativement faciles à suivre. L'analyste n'a que peu besoin d'un jugement indépendant. Par conséquent, les méthodes quantitatives obtiennent généralement des niveaux de fiabilité et de validité largement supérieurs à ceux des méthodes qualitatives. En revanche, l'interprétation des données qualitatives relève du jugement. Par conséquent, les méthodes qualitatives sont plus difficilement généralisables. Étant donné les avantages et limites de chaque catégorie, une démarche mixte, mêlant méthodes qualitatives et quantitatives, est souvent recommandée pour avoir un aperçu complet de la mise en œuvre et de l'efficacité du programme. Le tableau 3.4. présente les techniques courantes de collecte de données.

Avec le développement et l'expansion rapide des technologies de l'information et de la communication, il devient possible d'avoir recours à une gamme de plus en plus grande d'outils technologiques rendant plus accessible la collecte de données. Cela comprend l'utilisation de téléphones mobiles et autres appareils mobiles dans la mise en œuvre d'enquêtes, d'outils basés sur Internet, d'instruments de cartographie, et d'autres solutions multimédias.

[Ressource en ligne]

Aperçu des instruments de collecte de données fondés sur les technologies de l'information et de la communication

http://www.iyfnet.org/gpye-m&e-resource2

TABLEAU 3.4 Aperçu des méthodes de collecte de données

Méthode	Description	Utilisation	Avantages	Limites
Fichiers administratifs et de gestion	Documents qui fournissent des informations sur les processus de gestion de projet	Utiliser l'efficacité de la gestion de projets ou de l'exécution de la stratégie	• Fournir des informations sur un processus, informations difficiles à obtenir autrement	• Pour un programme donné, non généralisable • Dépend de la fiabilité des dossiers relatifs au système de gestion
Visites sur le terrain (observation et entretiens combinés)	Examen approfondi d'un site ou lieu spécifique	Assurer le suivi et comprendre le contexte	• Niveau de précision élevée • Accès aux données d'observation	• Pour un programme donné, non généralisable • Dépend fortement de l'accès aux lieux
Entretiens avec des informateurs clés	Méthode approfondie de collecte de données avec des personnes très informées	Obtenir des informations très détaillées sur une question en particulier ou sur plusieurs questions	• Niveau de précision élevée • Peut traiter de sujets non prévus • Assez souple pour étudier les problèmes en profondeur • Peut bien rendre plusieurs points de vue des parties prenantes	• •Pour un programme donné, non généralisable • La qualité varie fortement en fonction des compétences de celui qui pose les questions et de la mesure dans laquelle celui qui répond se sent à l'aise
Groupes de discussion	Méthode approfondie de collecte de données auprès de membres d'un groupe spécifique de la population averti (par exemple, les femmes, les jeunes, les personnes âgées, les travailleurs)	Obtenir des informations spécifiques et très détaillées sur le point de vue des parties prenantes sur une question en particulier ou sur plusieurs questions	• Mêmes avantages que pour l'entretien avec des informateurs clés • Permet l'interaction avec et entre les participants	• Pour un programme donné, non généralisable • La qualité dépend largement de la dynamique de groupe (par exemple, les participants peuvent être influencés par le modérateur ou des membres du groupe dominants) • Difficultés d'interprétation • Analyse qui prend du temps
Observation directe	Méthode de collecte de données par observation directe (par exemple, observation en classe). Les informations sont notées dans un dossier ou un journal)	Obtenir des données en milieu naturel	• Grande précision des données collectées par un observateur neutre • Fournit des informations sur le comportement réel et non le comportement décrit par la personne observée	• Non généralisable • Fort risque de déformation par l'observateur • Difficultés liées à l'interprétation et à l'encodage

Méthode	Description	Utilisation	Avantages	Limites
Examen des registres officiels	Documents officiels qui donnent des informations générales ou des données historiques sur certains phénomènes	Étudier les processus ou tendances/ données historiques sous-jacentes pour certains phénomènes	• Fournit des informations qu'il serait difficile d'obtenir autrement • Peu coûteux	• Éventuelles restrictions d'accès • Vérifier la validité et la fiabilité des données • Les données ne correspondent peut-être pas exactement à ce dont on a besoin.
Mini-enquêtes	Bref questionnaire/enquête qui collecte un ensemble de données limité	Obtenir des données quantitatives auprès d'un nombre limité de personnes ou sur un nombre limité de questions	• Plus rapide et moins coûteux que les enquêtes sur les ménages	• Portée limitée, par conséquent pas toujours représentative
Enquêtes sur les ménages	Large éventail de questions dont les réponses peuvent être codées de manière cohérente	Obtenir des informations auprès de nombreux participants sur leur situation socio-économique, des données démographiques, des modes de consommation, etc.	• Fournit des informations intéressantes et approfondies sur la population • Davantage généralisables que les mini-enquêtes • Peuvent être conçues pour collecter des données particulièrement intéressantes	• Chères • Pour être valables, elles doivent être menées par des personnes ayant des compétences particulières. • Difficile de convaincre les personnes de répondre à de longs questionnaires.
Enquêtes longitudinales	Étude longitudinale dans laquelle les variables sont mesurées avec les mêmes unités au fil du temps	Comme pour les enquêtes sur les ménages, avec un intérêt particulier pour la mesure des changements au fil du temps.	• Comme pour les enquêtes sur les ménages • Peut montrer la dynamique au fil du temps	• Comme pour les enquêtes sur les ménages • Parfois difficile de faire que les participants acceptent de répondre sur une longue période
Recensement	Enquête sur toute la population	Obtenir des données complètes sur une population donnée	• Généralisable • Généralement disponible auprès des sources officielles	• Coûteux • Prend du temps • Il est peu fréquent ; il date.

Sources : d'après Baker (2000); Creswell (2008).

[Conseil]

Utiliser des méthodes quantitatives quand :

- Il faut des données chiffrées généralisables pour convaincre les décideurs.
- Vous avez besoin d'informations statistiques représentatives sur la population cible, sa situation, ses comportements et ses attitudes.

Utiliser les méthodes qualitatives quand :

- Il faut répondre aux questions « comment » et « pourquoi », c'est-à-dire quand les données quantitatives doivent être expliquées par la motivation et les attitudes qui touchent les comportements.
- Les démarches participatives sont favorisées.

Les mécanismes de collecte de données sont plus ou moins adaptés aux différents niveaux de la chaîne de résultats. Les ressources et les indicateurs relatifs au processus s'appuieront principalement sur les dossiers de gestion et du projet qui montrent comment les ressources sont utilisées et les activités menées. L'observation directe et les visites sur le terrain peuvent fournir des données sur les indicateurs de résultats, par exemple le nombre de petites entreprises créées. Mesurer le résultat final exige souvent un mélange entre enquêtes formelles qui fournissent des informations quantitatives fiables et méthodes qualitatives, par exemple l'entretien avec les participants ou les groupes de discussion pour comprendre les mécanismes sous-jacents de l'échec ou de la réussite de certains effets et de la manière dont ils sont obtenus. Enfin, puisque les résultats au plus haut niveau portent généralement sur de grands changements dépassant le contrôle que l'on a sur le projet, les statistiques officielles peuvent être utiles pour de petites zones géographiques (par exemple les municipalités) et être ventilées par caractéristique sociodémographique.

Définir la fréquence et le moment de la collecte de données

L'intervalle entre les activités de suivi dépendra des objectifs du suivi. En règle générale, plus le niveau de la chaîne de résultats est élevé, moins il faut collecter fréquemment des données. Il est cependant plus difficile d'obtenir des informations précises.

Pour illustrer la meilleure fréquence de collecte des données qui soit, imaginons un programme de formation professionnelle trimestriel. Pour que la formation se déroule de manière efficace et efficiente, nous avons besoin d'informations sur la quantité de ressources utilisées (budget, temps du personnel, matériel, etc.) et sur l'exécution de nos activités (nombre d'heures de formation par semaine, nombre de participants, etc.). Ces informations sur nos ressources et activités devraient être recueillies assez régulièrement, disons toutes les deux semaines.

L'évaluation de notre résultat (nombre et composition du groupe des bénéficiaires de la formation) serait probablement régulière, disons mensuelle, bien que ces informations relèvent de données relatives à l'assiduité qui peuvent avoir été relevées chaque jour.

On ne saura clairement si la formation a eu un effet sur les résultats (connaissances des jeunes et capacités à trouver un emploi) qu'après la fin de la formation. Les effets à court terme, tels l'augmentation des connaissances, peuvent être mesurés à la fin de la formation, tandis que les effets qui se manifestent plus tard, par exemple la sécurité de l'emploi, ne seront mesurés que trois à six mois après l'intervention.

Enfin, les résultats au plus niveau, par exemple l'augmentation du revenu des ménages et les retombées positives ne se manifestent généralement pas avant une année (selon le marché du travail local l) et ne seront donc mesurés qu'à de longs intervalles.

[Conseil]

Le moment de la collecte de données devrait être fixé en fonction des réalités locales pour qu'il ne constitue pas un fardeau pour l'individu ou la famille. Les données ne devraient pas être collectées lorsque les jeunes ont des examens à l'école, par exemple, ou lorsque l'on a besoin de la main-d'œuvre des jeunes pendant certaines saisons agricoles.

Définir le responsable de la collecte des données

Il est important de définir clairement les responsabilités en matière de collecte des données. Ne pas le faire entraînera probablement un échec. Dans la pratique, plusieurs types de suivi incombent à différents acteurs, tant sur le terrain qu'au siège. Les personnes suivantes peuvent collecter des données :

- Administrateurs de programme
- Membres de l'équipe locale du projet ou responsables du suivi et de l'évaluation
- Partenaires d'exécution locaux (par exemple, les enseignants, les formateurs, les responsables des prêts)
- Bénéficiaires
- Autres parties prenantes (y compris les parents et les membres de la communauté)

- Recenseurs volontaires (par exemple, les étudiants à l'université)
- Consultants externes
- Instituts de sondage

Tout en définissant les responsables en matière de collecte des données, préciser ce qu'il advient des informations une fois recueillies. Intégrer au plan de collecte des données les procédures de stockage, de compilation et d'analyse des données pour garantir que les personnes qui en ont besoin y auront accès en temps voulu (voir ci-après le système de suivi et d'établissement de rapports).

Pour en savoir plus sur le suivi et l'évaluation participatifs, veuillez consulter Sabo Flores (2008), Powers et Tiffany (2006), et Gawler (2005).

Étape 3 : Articulation des risques et des hypothèses

Quels sont les principaux facteurs qui pourraient diminuer les effets potentiels du projet ? Quelles mesures peuvent être prises pour les atténuer ? Dans tout projet, il y a des facteurs incontrôlables qui auront des effets sur la réussite de notre intervention. Il peut notamment s'agir du temps, de la stabilité politique, de la situation en matière de sécurité locale et de l'appui des acteurs locaux. Une bonne compréhension de ces facteurs est essentielle à la conception du projet, ainsi qu'au suivi et à l'évaluation.

Définir les hypothèses pendant la phase de conception

Nous pouvons définir les hypothèses en réfléchissant aux facteurs essentiels à la réalisation de nos objectifs à chaque niveau de la chaîne de résultats et à ce qui pourrait les affecter (voir tableau 3.5.). Parfois, un premier groupe d'hypothèses peut avoir déjà été formulé dans la section *risque* de notre proposition de projet (Development Marketplace 2008).

[Conseil]

Lorsque de l'attribution des responsabilités en matière de recueil des informations et d'établissement des rapports, faites attention aux conflits d'intérêts. Par exemple, les enseignants ou les formateurs peuvent être tentés de mentir quant aux résultats obtenus (par exemple le nombre d'heures de formation) ou aux résultats finals (par exemple le nombre de jeunes qui ont amélioré leurs résultats aux tests ou trouver un emploi). Pour garantir la fiabilité des données, nous recommandons 1) d'utiliser des observateurs neutres pour garantir un suivi indépendant, et 2) de vérifier l'exactitude des informations fournies, au moins de façon sporadique, par des visites surprises sur site ou d'autres moyens.

Veuillez consulter la page http://www.povertyactionlab.org/evaluation/encouraging-teacher-attendance-through-monitoring-cameras-rural-udaipur-india pour voir comment le suivi photographique a rendu l'enseignant plus assidu et les visites de suivi moins nécessaires en Inde.

TABLEAU 3.5 Exemples d'hypothèses et de réponses dans le cadre du projet

Catégorie	Hypothèse	Sous notre contrôle ? Oui/Non
Ressource	• On peut trouver des formateurs qui souhaitent travailler dans le domaine du projet. • L'association des employeurs est prête à participer.	• Oui, mais pas encore embauchés. • Oui, protocole d'accord déjà signé.
Activité	• Électricité disponible sur le lieu de formation.	• Non, mais aucun problème ces derniers mois.
Résultat	• Les jeunes peuvent suivre la formation sans avoir à travailler pour aider leur famille.	• Non, mais des bons ont été distribués pour compenser l'absence de revenus.
Résultat final	• La formation est pertinente pour les besoins du marché du travail et est de grande qualité.	• Oui, des enquêtes sur les employeurs sont effectuées et les résultats des formateurs feront l'objet d'un suivi.
Résultats au plus haut niveau	• L'économie locale (notamment les prix du marché et les salaires) demeurent stables.	• Non, mais les prévisions sont bonnes.

Les hypothèses que nous ne contrôlons pas devraient être insérées dans la grille des résultats, au niveau qu'elles influencent. De manière générale, les ressources et les activités sont plus probablement sous le contrôle du projet que les résultats finals et les résultats au plus haut niveau.

Formuler des hypothèses irréalistes sur des éléments fondamentaux du programme peut gravement nuire à la réussite de l'intervention. Il faut donc toujours évité de le faire. Cela peut se produire quand :

- Nous surestimons les ressources dont nous disposons.

- Nous n'en savons pas suffisamment sur les bénéficiaires et le contexte local.
- Nous ne sommes pas en mesure d'évaluer correctement les facteurs de risques internes tels que l'insécurité ou l'opposition des autorités locales.

Source : d'après Development Marketplace 2008

Suivre les hypothèses pendant l'exécution du projet

Pour offrir un système d'alerte précoce sur des limitations potentielles et des solutions possibles, les hypothèses doivent être suivies de près. Cela nous permettra de savoir comment elles peuvent toucher la mise en œuvre et les résultats du projet, et ainsi nous aider à expliquer les écarts et à prendre les mesures correctives.

Mise en place d'un système de suivi et d'établissement de rapports

Planification

Une fois le cadre logistique complet avec indicateurs, instruments de collecte des données et hypothèses élaborés, les tâches suivantes vous permettront de vous préparer au suivi.

Concevoir les instruments nécessaires. Les données systématiquement collectées avec des instruments bien conçus permettront d'établir rapidement, et de manière fiable, les rapports. Les instruments devraient être testés auprès d'une population pertinente pendant l'élaboration et les résultats cet exercice pilote devraient orienter la conception des instruments suivants.

Établir des procédures pour protéger les jeunes. Même si les gouvernements nationaux ne les exigent pas toujours, les normes professionnelles prévoient que les activités de collecte de données soient administrées de manière à protéger les droits et intérêts des participants. La nature exacte de ces procédures est soumise aux exigences locales, mais les points suivants sont les minimums vivement recommandés :

- Créer des instruments et des procédures de formation de l'intervieweur qui garantissent l'anonymat des jeunes participants.
- Faire signer des formulaires de consentement éclairé incluant les détails du projet et les risques éventuels liés à la participation au projet. Ces formulaires expliquent également clairement les droits des participants, par exemple le droit d'abandonner le processus à tout moment. Il faut obtenir un consentement oral des personnes qui ne savent pas lire.
- Obtenir un consentement éclairé du parent ou tuteur des personnes qui n'ont pas atteint l'âge légal du consentement, des personnes atteintes d'un handicap mental et des autres populations vulnérables. Il ne faut pas collecter de données concernant une personne ne pouvant pas donner son consentement.

Pour des orientations plus précises, veuillez consulter la partie consacrée à la protection des personnes dans la note 7.

Collecter les données selon les méthodes choisies. Dans la mesure du possible, les processus existants, tels que l'enregistrement des participants ou de l'assiduité, devraient servir à alléger le fardeau que représente la collecte des données pour le personnel et les participants.

Constituer la base de données. Si les données collectées sont complexes, il peut s'avérer utile d'employer un administrateur de bases de données expérimenté chargé d'établir les codes et procédures qui permettent à plusieurs utilisateurs de rechercher les données et d'en extraire les résultats, après une petite formation. Plusieurs systèmes de bases de données conviennent et le projet doit choisir un logiciel qui permette de trouver un juste équilibre entre sophistication de l'analyse et facilité d'emploi.

[Conseil]

Veiller à ce que les instruments utilisés puissent enregistrer plusieurs types d'informations relatives aux participants (adresse, courriel, numéro de téléphone), aux amis et à la famille, qui pourront aider à localiser le jeune, très mobile, plus tard. L'utilisation de médias sociaux, par exemple Facebook, peut servir à communiquer avec les jeunes et à garder leur trace.

[Ressource en ligne]

Échantillons d'instruments d'enquête, dont certains contiennent des formulaires de consentement.

http://www.iyfnet.org/gpye-m&e-resource11

Compilation et analyse de l'information

Les méthodes de compilation et d'analyse des conclusions dépendent grandement des méthodes employées pour suivre un projet ou une intervention. Par conséquent, il faut décider très tôt dans le processus de conception comment utiliser les données de suivi. L'équipe de projet doit décider de la meilleure manière d'organiser ces données et de mener une analyse efficace et efficiente. Pour faciliter l'analyse et l'établissement de rapports dans de grands programmes, il est souhaitable de mettre en place d'un système de gestion de l'information reliant toutes les bases de données utilisées par les différentes unités du programme.

Pour les données qualitatives, il est souvent idéal (même si cela est difficile d'un point de vue logistique) d'utiliser un logiciel informatique d'analyse qualitative. Il existe plusieurs marques (par exemple Atlas.ti, NVivo, ou MaxQDA) qui fonctionnent toutes de manière similaire. Un logiciel d'analyse permet à un utilisateur d'importer tous les documents pertinents (par exemple les transcriptions d'entretiens ou des réunions des groupes de discussion, les documents du projet et les photographies) puis de leur appliquer un ensemble de codes prédéfinis. Selon le niveau de l'utilisateur, les codes peuvent être utilisés comme outil organisateur (qui regroupe tous les sujets semblables en provenance de différentes sources) ou d'analyse du lien entre les sujets. L'équipe devrait choisir le logiciel qui répond à ses besoins par rapport à l'expérience de ses membres et des coûts qu'il représente.

Pour les données quantitatives, lorsque les ressources le permettent, il est souvent mieux d'utiliser plusieurs systèmes, dont l'un devrait être une base de données relationnelle, telle que Microsoft Access. Les bases de données relationnelles permettent d'effectuer facilement des recherches et d'afficher les données selon plusieurs variables. Cependant, généralement, les analyses effectuées avec des bases de données relationnelles sont assez descriptives et offrent des mesures de tendance centrale (par exemple, moyenne, mode, médiane, écart-type). Si le projet l'exige et que les instruments sont conçus et administrés pour permettre une analyse plus poussée, le personnel chargé du suivi peut choisir d'utiliser un ensemble de logiciels tels que SPSS ou Stata. Outre ces ensembles de logiciels courants qui utilisent le disque dur d'un seul ordinateur, on utilise de plus en plus de systèmes d'analyse et de gestion des données qui permettent à une grande équipe de travailler ensemble au suivi et à l'analyse (« informatique en nuages »).

Apprentissage et prise de décisions

Le suivi ne sert pas à grand-chose si nous ne tirons pas d'enseignement et ne prenons aucune mesure après les conclusions de l'analyse. Se trouver en permanence dans un cycle d'action/réflexion permet de ne jamais oublier que les situations changent, que les besoins des bénéficiaires du projet peuvent changer et que les stratégies et activités du projet doivent être revues. Les organisations et les projets stagnent lorsqu'aucun enseignement n'est tiré ; un suivi rigoureux nous oblige à continuer d'apprendre (Shapiro 2003).

D'après Shapiro (2003), pour traduire les enseignements en actes, il faut :

- Se pencher sur les conséquences éventuelles de notre analyse sur notre programme.
- Énumérer les possibilités d'action.
- Examiner les options avec les parties prenantes internes et externes, parvenir à un consensus et obtenir un mandat d'action.
- Communiquer les ajustements et les plans au reste de l'organisation et, le cas échéant, à nos donateurs et bénéficiaires.
- Exécuter le plan.
- Suivre les effets.

[Définition]

Un **système de gestion de l'information** associe la technologie informatique, les personnes et les procédures en place pour recueillir, organiser et analyser des informations afin d'appuyer la prise de décisions. Il permet d'administrer de manière centralisée une grande quantité de données et de comparer les indicateurs par caractéristique des bénéficiaires et sur une période donnée.

En 2011, Youth Business International (YBI), un réseau de plus de 34 programmes indépendants axés sur l'entrepreneuriat des jeunes dans le monde, a entamé la mise en place d'un système de gestion des opérations en nuages à des fins de suivi. Ce système permet aux membres du réseau de suivre et d'analyser un large éventail d'indicateurs de réussite essentiels portant sur l'efficacité de l'organisation et les résultats finals. La qualité du portefeuille de prêts d'un membre et la réussite des affaires des entrepreneurs peuvent être évaluées au regard de facteurs tels que les caractéristiques sociodémographiques de l'entrepreneur, le parrainage et la formation existants, et les conditions du prêt. Cette plate-forme permet d'être plus précis et permet à l'équipe du réseau de compiler des informations en temps réel.

Établissement des rapports

Généralement, plus notre interlocuteur est haut placé dans l'organisation, moins nous avons besoin de fournir de détails et d'explications lorsque nous présentons nos conclusions. Présenter des messages clairs assortis de données agrégées et d'informations concises est généralement plus adapté à des interlocuteurs de haut niveau, principalement intéressés par la vue d'ensemble. Nous pouvons adapter le format de nos rapports aux interlocuteurs (voir tableau 3.6.).

TABLEAU 3.6 Adapter les rapports aux interlocuteurs

Interlocuteur	Format	Quand/Fréquence
Personnel du projet	Exposé oral et résumé écrit des statistiques lors des réunions d'équipe	Hebdomadaire
Équipe de gestion	Rapports écrits et exposé oral	Mensuel
Partenaires	Exposé oral et résumé écrit des statistiques	Mensuel
Donateurs	Dépend des exigences du donateur. En règle générale, de brefs rapports écrits soulignant les avancées du projet, les problèmes rencontrés, les résultats finals et les effets, l'efficacité de l'intervention/de la stratégie, etc.	Trimestriel/semestriel

Les données devraient toujours être données avec leur ligne de référence et les valeurs ciblées et présentées sous un « format simple, clair et facilement compréhensible » (Kusek et Rist 2004, p. 133). Les outils visuels, tels graphiques, diagrammes et cartes, peuvent être très utiles pour faire ressortir les données et messages essentiels.

Ressources

Les systèmes de suivi peuvent coûter cher. Outre les coûts fixes (matériel et logiciel informatiques, personnel), il y a aussi les coûts variables, notamment la formation des recenseurs locaux, le recrutement de consultants externes et la publication des conclusions (voir tableau 3.7.). Il est important de budgétiser correctement le système de suivi et d'évaluation et d'en tenir compte dans tout plan stratégique. Il arrive souvent que les administrateurs de programmes, lorsqu'ils réalisent le montant que cela représente, hésitent à octroyer d'importantes ressources à un système de suivi et d'évaluation, car cela semble se faire aux dépens des activités d'intervention. Toutefois, sans système de suivi adapté, un programme risque de ne pas obtenir les meilleurs résultats ou d'échouer, sans que l'on soit conscient de ces éventualités. En outre, sans suivi nous ne pourrons peut-être pas saisir les possibilités de large succès. En fin de compte, les systèmes de suivi sont essentiels à la gestion de projets et constituent un élément fondamental de toute intervention.

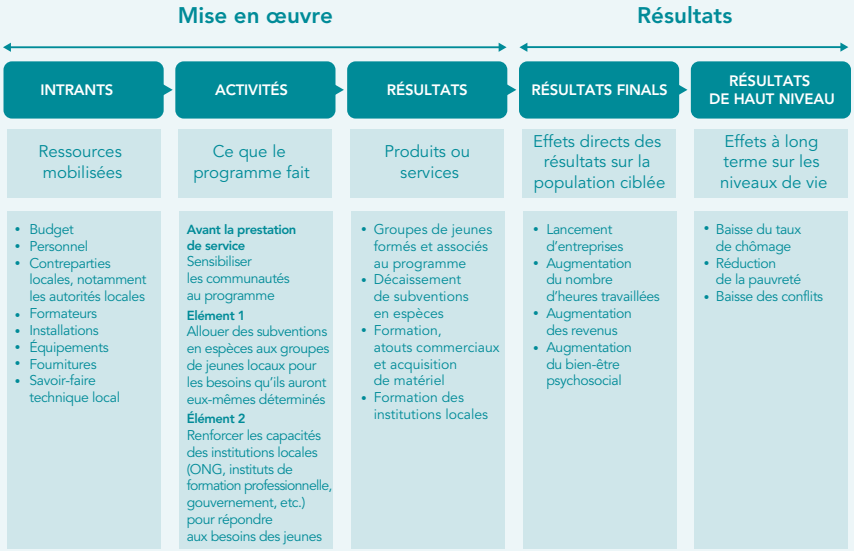
TABLEAU 3.7 Éléments habituels d'un budget pour le suivi

Coûts fixes	
Dépenses de personnel	• Siège : Temps que le coordonnateur chargé du suivi et de l'évaluation consacre à la gestion du système de suivi et d'évaluation, en pourcentage. Selon la taille du projet, cela peut aller de 10 à 100 %. • Localement : Généralement 50 à 100 % du temps d'un responsable du suivi et de l'évaluation au niveau local pour gérer l'exécution des activités de suivi et d'évaluation, et personnel d'exécution d'appui.
Équipement	Ordinateurs, enregistreurs audio, caméras, etc.
Logiciel	Licences pour des outils d'analyse quantitative et qualitative
Coûts variables	
Formation	Renforcement des capacités du personnel, des recenseurs, des membres des communautés, etc.
Voyage	Voyage du personnel du Siège sur le terrain pour des vérifications périodiques et une aide technique. Déplacements au niveau local pour veiller à une mise en œuvre normalisée des activités de suivi et d'évaluation.
Collecte et analyse de données	Recourir à des tiers, par exemple les instituts de sondage.
Consultants	Recourir à des experts externes pour certaines tâches.
Impression	Instruments, rapports, etc.

Points essentiels

1. Chaque intervention doit être dotée d'un système de suivi solide pour pouvoir contrôler de manière continue la mise en œuvre et les résultats, que le projet fasse l'objet d'une évaluation ou pas.
2. Les administrateurs du programme et les principales parties prenantes doivent conjointement élaborer une chaîne de résultats pour spécifier clairement la logique de l'intervention et définir les indicateurs clés, les mécanismes de collecte des données et les hypothèses.
3. Le système de suivi fournit des informations en continu sur l'orientation, le rythme et l'ampleur du changement. Il permet également de repérer des événements imprévus dans le projet ou l'environnement dans lequel il se déroule. C'est grâce à cela que l'on peut savoir si une intervention se déroule dans la direction souhaitée ; le suivi de qualité est indispensable à une gestion de projet efficace.
4. Les données du suivi sont descriptives et n'expliquent pas nécessairement pourquoi et comment certains changements surviennent. Elles ne donnent pas les éléments qui permettent d'attribuer les changements à l'intervention, c'est-à-dire qu'elles ne prouvent pas que les changements se produisent du fait de notre programme.

Étude de cas du NUSAF : système de suivi


Mise en œuvre
Résultats
INTRANTS
ACTIVITÉS
RÉSULTATS
RÉSULTATS FINALS
RÉSULTATS DE HAUT NIVEAU
Ressources mobilisées
Ce que le programme fait
Produits ou services
Effets directs des résultats sur la population ciblée
Effets à long terme sur les niveaux de vie
• Budget
• Personnel
• Contreparties locales, notamment les autorités locales
• Formateurs
• Installations
• Équipements
• Fournitures
• Savoir-faire technique local
Avant la prestation de service
Sensibiliser les communautés au programme
Élément 1
Allouer des subventions en espèces aux groupes de jeunes locaux pour les besoins qu'ils auront eux-mêmes déterminés
Élément 2
Renforcer les capacités des institutions locales (ONG, instituts de formation professionnelle, gouvernement, etc.) pour répondre aux besoins des jeunes
• Groupes de jeunes formés et associés au programme
• Décaissement de subventions en espèces
• Formation, atouts commerciaux et acquisition de matériel
• Formation des institutions locales
• Lancement d'entreprises
• Augmentation du nombre d'heures travaillées
• Augmentation des revenus
• Augmentation du bien-être psychosocial
• Baisse du taux de chômage
• Réduction de la pauvreté
• Baisse des conflits

Pour asseoir l'interprétation des résultats de l'évaluation d'impact, il était essentiel pour le programme NUSAF de disposer de bonnes informations quant à la mise en œuvre du programme visant à améliorer les perspectives d'emploi des jeunes comme souhaité. Le NUSAF a donc utilisé des outils quantitatifs et qualitatifs pour suivre les activités et les résultats. Par exemple, puisque des subventions en espèces étaient allouées aux groupes de jeunes par les autorités centrales, on demandait aux jeunes s'ils avaient reçu l'argent. Ces informations étaient ensuite comparées avec les registres du gouvernement.

Le programme a aussi essayé de comprendre la distribution et l'utilisation de l'argent au sein du groupe. Comme l'argent devait servir à la formation, au matériel et aux instruments, le NUSAF a suivi le niveau d'assiduité, le nombre et la valeur des instruments et du matériel, et regardé si les jeunes avaient lancé une entreprise et si cette dernière fonctionnait toujours.

Même si ces informations n'ont pas fourni de réponses quant aux effets du programme, elles ont aidé les responsables du programme, le personnel de suivi et les évaluateurs à comprendre si le programme était exécuté comme prévu et comment il pouvait toucher les participants. Cela a aussi aidé à analyser les résultats de l'évaluation, par exemple en expliquant pourquoi certains participants avaient plus ou moins bénéficié du programme que d'autres.

Source : Blattman, Fiala et Martinez (2011).

À lire

Donor Committee for Enterprise Development. 2010. *The DCED Standard for Measuring Achievements in Private Sector Development. Control Points and Compliance Criteria.* Version V. http://www.enterprise-development.org/page/measuring-and-reporting-results

Fondation W. K. Kellogg. 2004. *Logic Model Development Guide.* Battle Creek, MI: W. K. Kellogg Foundation. http://www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx

Kusek, J. Z. et Rist, R. C, 2004, *Dix étapes pour mettre en place un système de suivi et d'évaluation axé sur les résultats : Un guide pour les praticiens du développement*, Washington, D.C., La Banque mondiale, voir chapitres 2 à 6, http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/11/16/000310607_20061116133234/Rendered/PDF/296720FRENCH0T1d1OFFICIAL0USE0ONLY1.pdf

Taylor-Powell, E., Jones, L. et Henert, E. 2003. *Enhancing Program Performance with Logic Models.* University of Wisconsin–Extension, Program Development and Evaluation. http://www1.uwex.edu/ces/lmcourse/

Notes

NOTE 4 : Choisir le bon type d'évaluation

Les plus graves erreurs ne sont pas le résultat de réponses erronées.
Ce qui est véritablement dangereux,
c'est de poser la mauvaise question.

—Peter Drucker

Bien qu'il soit essentiel de mettre en place un bon système de surveillance pour savoir si notre intervention va dans le sens voulu, cela ne permet pas nécessairement de savoir comment et pourquoi des changements surviennent, ni de prouver que tout changement observé dans les effets d'une action sont le résultat de notre intervention. Pour compléter les informations que nous obtenons au moyen de notre système de surveillance, nous avons besoin d'évaluations. Les évaluations sont des appréciations périodiques destinées à déterminer la pertinence, l'efficience, l'efficacité, l'impact et la durabilité de notre intervention. Le type d'évaluation le mieux adapté à notre projet dépendra avant tout de nos besoins en termes d'informations. Par conséquent, la première étape dans toute procédure d'évaluation est de définir ce que nous voulons apprendre. Ces objectifs d'apprentissage ainsi que le contexte dans lequel se déroule notre intervention nous permettront de déterminer le type d'évaluation qui convient le mieux dans le cadre de notre programme.

Quel est le but de l'évaluation ?

Avant même de décider s'il est nécessaire de mener une évaluation et quel type d'évaluation il faudrait choisir, il est crucial de définir clairement ce que nous voulons tirer de l'évaluation. *Quelle décision sera fondée sur l'évaluation et quels types d'information faut-il recueillir pour prendre cette décision ?* Voulons-nous en savoir davantage sur la manière dont sont effectivement mis en œuvre nos programmes, savoir si nos programmes remplissent les objectifs ou si nos bénéficiaires ont réellement plus de moyens financiers à la suite de notre intervention ? En tant que responsables de programmes et évaluateurs, nous devons tout d'abord définir nos objectifs d'apprentissage et formuler nos questions, puis choisir l'outil d'évaluation le plus approprié pour obtenir les informations voulues (Karlan 2009).

En termes généraux, les évaluations permettent de répondre à trois types de questions (Imas et Rist 2009) :

- **Les questions descriptives** permettent de décrire les processus, les conditions, les relations organisationnelles et les points de vue des acteurs (*Où en est notre projet* ?).
- **Les questions normatives** permettent de comparer les activités menées à celles qui devraient être menées. Elles permettent de comparer la situation actuelle avec les objectifs et les buts spécifiques qui ont été fixés (*Notre projet a-t-il été mis en œuvre comme prévu ? Les résultats obtenus étaient-ils prévus ?*).
- **Les questions sur les causes et les effets** permettent d'examiner les résultats obtenus et d'essayer de mesurer les changements apportés par une intervention. Elles permettent de savoir si les objectifs atteints résultent de notre projet (*Quel est l'impact ou quels sont les effets de causalité de notre programme sur les résultats d'intérêt ?*).

Les questions, parmi celles-là, que nous devrions nous poser ne dépendent finalement que de nous, selon l'intervention menée.

Organiser nos questions. Dans la pratique, il se peut que nous ayons beaucoup de questions à poser, relevant de toutes les catégories. L'examen de notre chaîne de résultats est un bon moyen d'organiser toutes les questions d'évaluation possibles (voir tableau 4.1.). En fait, si un bon système de surveillance est en place (voir note 3), il devrait y avoir un consensus autour de notre logique de projet en termes de mise en œuvre et de résultats, ce qui en retour facilite l'identification des objectifs d'apprentissage essentiels au cours de toutes les étapes de l'intervention. Les questions descriptives et normatives peuvent se rapporter à tous les niveaux de la chaîne de résultats ; mais les questions sur les causes et les effets se rapportent spécifiquement aux résultats obtenus, y compris les résultats au plus haut niveau.

TABLEAU 4.1 Exemples de questions d'évaluation

	Intrants	Activités	Résultats	Résultats finals	Résultats au plus haut niveau
Descriptives	• Le coût du programme est-il comparable à celui d'interventions analogues ? • Quelles sont les qualifications des prestataires de services ? • Quelles sont les autres interventions en cours ?	• Les jeunes connaissent-ils le programme et remplissent-ils les conditions requises pour y participer ? • Quels sont les mécanismes utilisés pour la mise en œuvre ? • Dans quelle mesure la mise en œuvre du programme diffère-t-elle selon le contexte ?	• Combien de jeunes participent-ils (par âge, sexe, etc.) ? • Qui abandonne ? • Quels sont les services les plus utilisés ?	• Les participants sont-ils satisfaits du programme ? • Y a-t-il des changements observables dans les connaissances des participants, leurs attitudes, etc. ? • Combien de participants au programme trouvent-ils un emploi dans les 3 mois ?	• Le chômage parmi les jeunes augmente-t-il ou baisse-t-il au niveau local ? • Les revenus des ménages évoluent-ils ?
Normatives	• Dépensons-nous autant que prévu ? • Les ressources humaines et financières sont-elles suffisantes ? • Le programme duplique-t-il d'autres activités ?	• La procédure de sélection des participants est-elle impartiale et équitable ? • La mise en œuvre du programme est-elle retardée ? • Les manuels opérationnels sont-ils suivis ?	• L'équilibre hommes-femmes souhaité entre les participants est-il atteint ? • L'objectif visant à former 5 000 jeunes par an sera-t-il atteint ?	• Les revenus des participants augmentent-ils de 20 % comme prévu ? • 80% des bénéficiaires trouvent-ils un emploi dans les 3 mois suivant leur qualification, comme prévu ? • Quels sont les éventuels effets positifs ou négatifs qui n'étaient pas prévus ?	• Davantage de ménages deviennent-ils autonomes financièrement ? • Davantage de ménages parviennent-ils à la sécurité alimentaire?
Sur les causes et les effets	Sans objet	Sans objet	Sans objet	• À la suite de la formation à l'emploi, les participants ont-ils un emploi mieux rémunéré ? • La mise en place de stages pratiques augmente-t-elle l'efficacité de la formation technique proposée ? • Le programme touche-t-il les garçons et les filles de manière différente ?	• Le projet contribue-t-il à réduire la pauvreté dans la zone ciblée ? • Quels sont les autres effets de cette intervention sur les conditions de vie de la communauté dans son ensemble ?

Le lien entre questions d'évaluation et critères d'évaluation. Penser aux critères communs d'évaluation tels qu'ils ont été définis à l'origine par l'Organisation de coopération et de développement économiques (OCDE) est un autre moyen d'envisager les questions d'évaluation. Comme il a déjà été mentionné, les évaluations sont des appréciations périodiques destinées à déterminer la *pertinence*, l'*efficience*, l'*efficacité*, l'*impact* et la *durabilité* de notre intervention (OCDE 1991). En y regardant de plus près, nous voyons que la *pertinence*, l'*efficience* et l'*efficacité* se rapportent essentiellement aux questions normatives, alors que l'*impact* a trait à la causalité. Les questions relatives à la *durabilité* peuvent être normatives (est-il probable que l'intervention se poursuive après la fin du financement par les donateurs ?) ou porter sur les causes et les effets (les effets observés sont-ils durables au fil du temps ?) Aucune de ces questions n'est purement descriptive, bien que les questions normatives comprennent par essence des éléments descriptifs. Le tableau 4.2. reprend chaque critère correspondant au type de question d'évaluation.

TABLEAU 4.2 Le lien entre critères d'évaluation et questions d'évaluation

Critères	Description	Détails	Type de question d'évaluation
Pertinence	Les objectifs correspondent-ils aux problèmes ou aux besoins visés ?	• Dans quelle mesure les objectifs du programme sont-ils encore valables ? • Les activités menées dans le cadre du programme et les résultats obtenus sont-ils cohérents avec la réalisation de l'ensemble des objectifs fixés ?	Normative
Efficience	Le projet est-il exécuté dans les délais et est-il efficace par rapport à son coût ?	• Le programme ou le projet sont-ils mis en œuvre de la manière la plus rentable qui soit ? • Quels sont les coûts par résultat/bénéficiaire et sont-ils comparables à ceux d'interventions analogues ?	Normative
Efficacité	Dans quelle mesure l'intervention remplit-elle ses objectifs ?	• Dans quelle mesure les résultats escomptés ont-ils été obtenus ? • Quels sont les principaux facteurs qui influencent la réalisation ou la non-réalisation des objectifs ?	Normative
Impact	Quels sont les changements positifs et négatifs produits par l'intervention ?	• Quels sont les résultats au plus haut niveau produits par le programme ou le projet ? • Quel changement concret l'activité a-t-elle apporté dans la vie des bénéficiaires ?	Sur les causes et les effets
Durabilité	Y a-t-il des bénéfices durables une fois que l'intervention a été menée à terme ?	• Dans quelle mesure les bénéfices d'un projet perdurent-ils après la fin du financement par les donateurs ? • Quels sont les principaux facteurs qui influencent le fait qu'un projet est durable ou non ?	Normative ou sur les causes et les effets

Source : d'après l'OECD (n.d.)

Classer nos questions par priorité. Aucun type de question ne peut en remplacer un autre, bien que les questions normatives comprennent généralement des questions descriptives ou sont fondées sur ce type de question. Tous les types de questions portent sur des aspects différents du projet et apportent un type d'information différent, qui peut être utile. Toutefois, si nous voulons nous concentrer sur les résultats, alors les questions sur les causes et les effets sont particulièrement intéressantes. En fait, si nous voulons déterminer des interventions prometteuses pour améliorer les moyens de subsistance des jeunes et montrer quels sont les effets réels de notre intervention, alors les questions sur les causes et les effets devraient constituer une partie – si ce n'est une priorité – de nos objectifs d'apprentissage dans le cadre notre programme.

Chacun de ces trois types de questions – descriptives, normatives et portant sur les causes et les effets – conduit à prendre en considération des éléments différents pour déterminer le genre d'évaluation à mettre en place. Les responsables de programmes et les évaluateurs peuvent estimer qu'une question relève d'un des trois types, puis examiner les implications de chaque type de question sur l'élaboration d'une procédure d'évaluation. Ainsi, en choisissant une série de questions d'évaluation, nous définissons l'éventail d'outils de surveillance et d'évaluation appropriés qui nous permettront d'y répondre (GAO 1991).

[Conseil]

Assurez-vous d'identifier le public visé par l'évaluation et de déterminer ce que ce public veut savoir. Certaines évaluations peuvent être requises au sein de l'organisation par les personnes chargées de mettre en œuvre le programme ou par la direction. Les donateurs ou les décideurs peuvent en exiger d'autres. Les besoins internes et externes en matière d'informations peuvent être différents, ce qui suppose des questions d'évaluation différentes. Il est donc crucial d'associer les acteurs à l'élaboration des questions d'évaluation et à leur classement par priorité.

Source : d'après Rubio (2011).

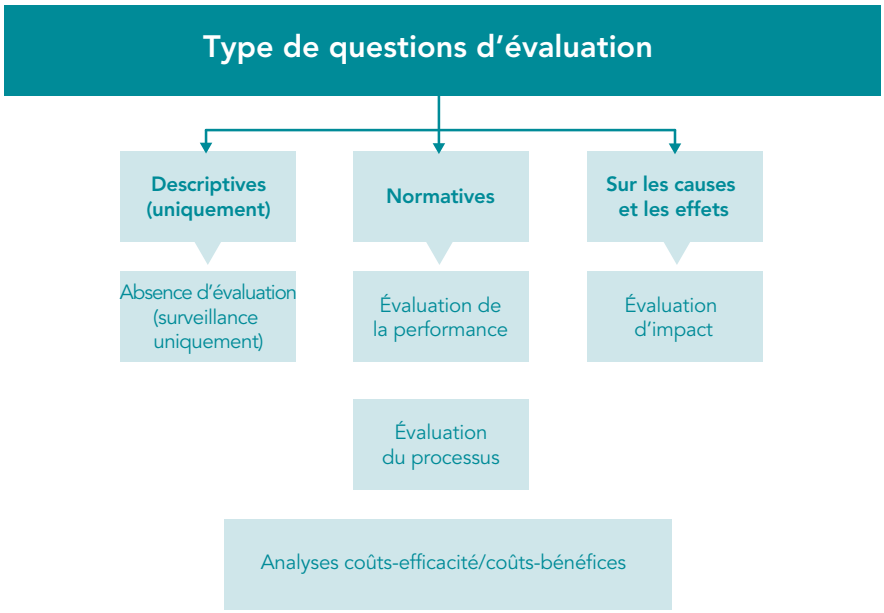
Relier les questions d'évaluation à la procédure d'évaluation

Il n'y a pas de modèle d'évaluation unique qui conviendrait à tous les cas. En définitive, le choix de l'évaluation devrait dépendre des questions susmentionnées et non pas de nos préférences méthodologiques ou de celles de l'évaluateur interne ou externe. Cela peut sembler évident, mais ce n'est pas toujours le cas.

La figure 4.1. nous donne un aperçu des procédures d'évaluation possibles selon le type de question que nous voulons privilégier[2].

2 Il existe d'autres types d'évaluation axés sur d'autres niveaux de la fourniture d'aide (comprenant les secteurs, les thèmes et l'efficacité de l'aide apportée) qui ne sont pas pris en compte dans cette note. Cette note se limite à l'évaluation de projets et de programmes.

FIGURE 4.1 Des questions d'évaluation à la procédure d'évaluation



Absence d'évaluation

Si un responsable de projet a seulement besoin d'informations descriptives sur l'intervention, par exemple parce que la réalisation du projet en est à un stade très précoce et que l'objectif est d'obtenir des informations générales sur la façon dont le programme est mis en œuvre, alors il n'est pas forcément nécessaire de mener une évaluation complète. Dans ce cas, les informations obtenues au moyen de la surveillance peuvent suffire. Évidemment, il faut pour cela qu'un système de surveillance efficace soit en place, avec une chaîne de résultats clairement définie, ainsi que des indicateurs, des outils de collecte de données, etc. (voir note 3). Si un tel système existe, il devrait être relativement facile de disposer d'informations descriptives sur le programme.

Évaluation de la performance

Les évaluations de la performance permettent d'évaluer dans quelle mesure les objectifs d'un programme ont été bien définis (voir les critères dans la note 2), ainsi que les progrès accomplis dans la réalisation de ces objectifs (Rubio 2011). Elles permettent également de savoir si le cadre de résultats mis en place est approprié ; c'est-à-dire, s'il y a des incohérences entre les ressources, les activités et les objectifs, et si les priorités ou les délais devraient être adaptés en vue d'une meilleure réalisation des objectifs fixés. De telles évaluations peuvent être conduites à tous les stades de la mise en œuvre, mais elles sont plus particulièrement courantes lors des examens à mi-parcours (lorsqu'il faut recueillir des informations pour la gestion du projet) ou à la fin du projet (lorsqu'elles portent sur le compte rendu et les enseignements tirés en vue d'interventions futures). Généralement conduites par un évaluateur indépendant, les évaluations de la performance peuvent être mises en œuvre relativement vite et pour un coût modéré, car elles sont principalement fondées sur une recherche documentaire et des entretiens sélectionnés. Il arrive toutefois qu'elles soient fondées sur une collecte de données plus vaste, comme la comparaison avant et après des résultats des participants ou d'autres outils de mesure de la qualité. Bien qu'utiles à des fins d'évaluation générale de la qualité, les évaluations de la performance ne visent pas à mesurer rigoureusement les effets d'un programme, ni permettent de savoir avec certitude si les changements observés sont dus à une intervention donnée.

Évaluation du processus

À la différence des évaluations de la performance, qui portent essentiellement sur la réalisation des objectifs, les évaluations du processus visent à comprendre globalement comment un programme se déroule et à évaluer dans quelle mesure un projet est mis en œuvre correctement. Elles permettent de déterminer s'il y a des divergences entre les activités prévues et réalisées et les résultats, et de mieux comprendre les raisons de ces divergences. Fondées sur des informations descriptives, telles que les activités offertes et les personnes participant au programme (ou celles n'y participant pas), elles permettent aussi de recenser les moyens d'améliorer la qualité des services offerts. Une évaluation de la performance peut être conduite à des étapes particulières en tant que système d'alerte rapide ou lorsque des problèmes, comme des retards dans la mise en œuvre ou l'insatisfaction des bénéficiaires, ont été relevés lors des procédures de surveillance ordinaires (Banque mondiale 2002). Les évaluations du processus tendent à se baser sur un mélange d'outils de mesure de la quantité et de la qualité, notamment les entretiens avec des informateurs clés, les enquêtes de satisfaction auprès des utilisateurs, l'observation directe et les groupes cibles.

Évaluation d'impact

[Définition]

On entend par **contrefactuel** les résultats auxquels seraient parvenus les participants du programme en absence d'intervention. Le contrefactuel apporte une réponse à la question : *Que serait-il arrivé au bénéficiaire si le programme n'avait pas été mis en place ?*

Les évaluations d'impact sont le seul type d'évaluation qui permette de répondre spécifiquement aux questions sur les causes et les effets. Avec ces questions, il nous faut déterminer non seulement si les résultats escomptés ont été obtenus mais également si les résultats ont été obtenus *parce que le programme avait été mis en œuvre*. Comme Gertler et al. (2011, p. 4) le font remarquer, cet accent mis sur la causalité et l'attribution des effets est la caractéristique des évaluations d'impact et détermine le type de méthodologies qui peuvent être utilisées (la note 6 nous donne un aperçu des outils appropriés). Pour estimer les effets de causalité d'un programme sur les résultats escomptés, toute méthode choisie doit permettre d'estimer ce qu'il est convenu d'appeler le *contrefactuel*, c'est-à-dire ce qui serait arrivé aux participants du programme en absence d'intervention. Pour ce faire, il faut trouver un groupe témoin, à savoir un groupe de personnes qui, en absence d'intervention, serait parvenu à des résultats analogues à ceux des bénéficiaires du programme (Duflo, Glennerster et Kremer 2006). C'est ce qui rend les évaluations d'impact différentes de toutes les autres évaluations. De ce fait, elles nécessitent généralement plus de temps et de compétences dans le domaine des analyses quantitatives, et elles sont souvent plus coûteuses que les évaluations d'autres types. Étant donné les informations qu'elles permettent de recueillir, les évaluations d'impact sont particulièrement utiles pour apporter une réponse aux questions d'ordre stratégique, allant du renforcement des interventions jugées efficaces à la réduction des projets peu prometteurs (Rubio 2011). En outre, elles contribuent à enrichir le niveau de connaissances générales sur la relative efficacité des différents types d'interventions visant à renforcer les moyens de subsistance et nous aident à mieux comprendre quelles options, lors de l'élaboration des programmes (dosage, moyen d'exécution, etc.), sont les plus importantes dans une catégorie spécifique de programmes.

Analyse coûts-efficacité et coûts-bénéfices

Les évaluations des coûts-efficacité et des coûts-bénéfices permettent de mesurer les coûts monétaires et non monétaires du programme et de les comparer avec les utilisations autres des mêmes ressources et les bénéfices produits par le programme (Baker 2000). L'*analyse coûts-efficacité* permet d'évaluer les coûts par rapport aux résultats obtenus, ou aux résultats finals (par exemple, 300 dollars des États-Unis par jeune formé, 500 dollars par emploi créé) et de comparer ces coûts à ceux d'interventions semblables menées par notre organisation ou d'autres. Elle permet donc de savoir combien de résultats ont été obtenus pour chaque dollar dépensé (question descriptive) et s'il y a un écart par rapport à nos attentes (question normative). L'*analyse coûts-bénéfices* permet quant à elle de

mesurer la totalité des coûts prévus par rapport à l'ensemble des bénéfices escomptés (résultats finals) d'une intervention –les coûts et les bénéfices sont généralement exprimés en termes monétaires. Par exemple, si notre programme visait à aider 500 jeunes à trouver un emploi et à le conserver ou à créer une petite entreprise viable : 1) nous estimerions la totalité des avantages en termes de revenus plus élevés, de meilleure santé, de taux de criminalité plus faible, etc. ; et 2) nous comparerions ces avantages au coût total de l'intervention. Étant donné que l'analyse coûts-bénéfices porte sur la valeur des bénéfices perçus, elle doit être fondée sur une estimation crédible de la mesure dans laquelle le programme a influencé les résultats escomptés, ce qui la rend très utile en combinaison avec les évaluations d'impact (pour une description plus détaillée, voir la note 8). Des liens vers des exemples de types d'évaluations mentionnés plus haut sont fournis dans l'encadré 4.1.

ENCADRÉ 4.1 Exemples d'évaluations par type

Évaluations de la performance

- Human Sciences Research Council. 2007. *Mid-term Review of the Expanded Public Works Programme: Synthesis Report*. Pretoria : Southern Africa Labour and Development Research Unit, University of Cape Town ; Rutgers School of Law ; et ITT (Royaume-Uni). http://www.hsrc.ac.za/research/output/outputDocuments/5465_Hemson_MidtermreviewofEPWPsynthesisreport.pdf
- Education and Employment Alliance. 2010. *An Evaluation of Partnerships in Support of Youth Employability: Global Report*. http://www.iyfnet.org/document/1436

Évaluations du processus

- Miller, E., et MacGillivray, L. 2002. *Youth Offender Demonstration Project Process Evaluation*. Chapel Hill : Research and Evaluation Associates Inc. http://wdr.doleta.gov/opr/fulltext/YODP_final.pdf
- The Lewin Group, Inc. 2003. *Evaluation Design for the Ticket to Work Program – Preliminary Process Evaluation*. http://www.lewin.com/content/publications/2526.pdf

Évaluations d'impact

- Attanasio, O., Kugler, A. et Meghir, C. 2009. « Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial. », IZA Discussion Paper No. 4251. Bonn, IZA. http://ssrn.com/abstract=1426738
- Mensch, B., Grant, M., Sebastian, M., Hewett, P., et Huntington, D. 2004. "The Effect of a Livelihoods Intervention in an Urban Slum in India: Do Vocational Counseling and Training Alter the Attitudes and Behavior of Adolescent Girls?" Working Paper No. 124, New York, Conseil de population, http://www.popcouncil.org/pdfs/wp/194.pdf

Analyses coûts-efficacité et coûts-bénéfices

- Elias, V., Nuñez, F., Cossa, R., et Bravo, D. 2004. *An Econometric Cost-benefit Analysis of Argentina's Youth Training Program*. Washington, DC: IDB. http://www.iadb.org/res/publications/pubfiles/pubR-482.pdf
- Jastrzab, J., Masker, J., Blomquist, J., et Orr, L. 1996. *Evaluation of National and Community Service Programs—Impacts of Service: Final Report on the Evaluation of American Conservation and Youth Service Corps*. Bethesda, MD: Abt Associates Inc. http://www.abtassociates.com/reports/ccc_youth_0596.pdf (N. B. : évaluation d'impact et analyse coûts-bénéfices combinées)

Le contexte dans lequel se déroule notre programme correspond-il au type d'évaluation souhaité ?

Comme l'a noté le GAO (1991, p. 15), c'est une chose de se mettre d'accord sur les questions prioritaires et de choisir une procédure d'évaluation ; c'en est une autre de savoir s'il est possible d'apporter une réponse à ces questions et, si tel est le cas, à quel prix en termes d'argent, de personnel et de temps.

Après avoir formulé les bonnes questions et déterminé un type d'évaluation, nous devons analyser le contexte dans lequel se déroule l'intervention afin de comprendre quelle évaluation peut être conduite au vu des contraintes existantes.

Le choix du moment

Les questions visant à déterminer *quel type* d'information est nécessaire sont étroitement liées à la question tendant à savoir *quand* les résultats de l'évaluation doivent être disponibles. Le fait de savoir quand ils doivent être disponibles permet de déterminer quand les informations doivent être recueillies.

Quand estime-t-on qu'il est nécessaire de conduire une évaluation ?

Prévoir bien à l'avance l'évaluation donne plus de flexibilité pour choisir un outil adéquat. Par exemple, beaucoup de méthodes d'évaluation d'impact doivent être planifiées avant même le début de la mise en œuvre du programme. Dans l'idéal, la planification d'une évaluation devrait faire partie de la planification du programme (on parle d'évaluation prospective). Dans bien des cas, cependant, les besoins en matière d'informations peuvent apparaître soudainement, par exemple à la suite de problèmes soudains sur le terrain ou d'une demande émanant d'un donateur. De la même façon, les contraintes liées au déroulement du programme, telles que la mise en œuvre rapide d'un programme afin de décaisser des fonds, peuvent dicter le calendrier d'évaluation. Si ces contraintes sont inévitables dans la vie réelle, elles réduisent de fait les possibilités d'évaluation qui nous offertes dans certaines circonstances.

À quel stade de la mise en œuvre du programme est-il nécessaire de recueillir des informations ?

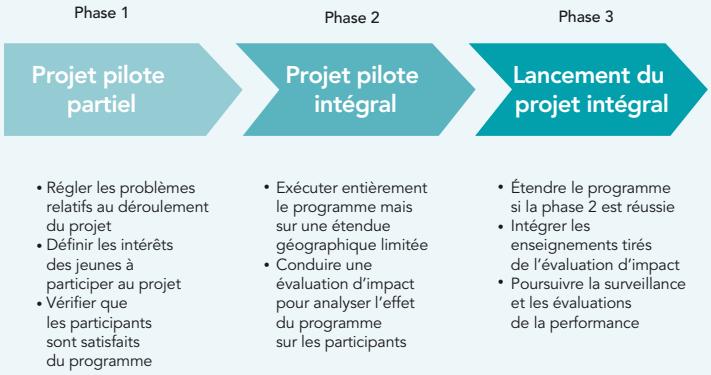
Les besoins en informations varient en fonction du stade de la mise en œuvre du programme. Par exemple, il peut être nécessaire de procéder à une analyse coûts-efficacité dans le cas d'un programme qui vient d'être planifié afin de déterminer s'il faut le mettre en œuvre ou non. Autrement, dans le cas d'une intervention venant d'être lancée, nous pouvons avoir besoin de savoir dans quelle mesure les procédures sont bien respectées, s'il faut faire des ajustements pour garantir la bonne réalisation du programme (Rubio 2011). Bien souvent, ces besoins en matière d'informations peuvent être évalués avant même le commencement du programme, de même que le moment approximatif de l'évaluation.

Combien de temps l'évaluation prend-elle ?

Le temps que prend une évaluation dépend en partie des méthodes utilisées pour collecter et analyser les données, ce qui diffère en fonction du type d'évaluation, ainsi que de l'ampleur et de la profondeur de l'étude, ce qui diffère pour chaque type d'évaluation. En général, on estime que les évaluations de la performance pures peuvent prendre entre un et trois mois, car elles sont principalement fondées sur une recherche documentaire et un nombre limité d'entretiens. Les évaluations du processus, en revanche, peuvent varier considérablement en termes d'envergure. Elles peuvent être aussi rapides que les évaluations de la performance mais prendre jusqu'à six mois ou plus si des processus complexes sont analysés. De toutes les évaluations, ce sont les évaluations d'impact qui tendent à prendre le plus de temps (de six mois à deux ans ou plus), car il faut bien planifier la méthode employée, et il peut être nécessaire d'entreprendre de nouvelles collectes de données. Les analyses coûts-bénéfices peuvent prendre moins d'un mois si toutes les données nécessaires sont disponibles. S'il faut d'abord recueillir les informations, elles peuvent prendre beaucoup plus de temps.

L'encadré 4.2. illustre à quel moment, dans les phases d'exécution d'un projet, il est préférable de mener certaines procédures d'évaluation.

ENCADRÉ 4.2 Phases d'exécution d'un programme et stratégies d'évaluation appropriées


Phase 1
Projet pilote partiel
• Régler les problèmes relatifs au déroulement du projet
• Définir les intérêts des jeunes à participer au projet
• Vérifier que les participants sont satisfaits du programme
Phase 2
Projet pilote intégral
• Exécuter entièrement le programme mais sur une étendue géographique limitée
• Conduire une évaluation d'impact pour analyser l'effet du programme sur les participants
Phase 3
Lancement du projet intégral
• Étendre le programme si la phase 2 est réussie
• Intégrer les enseignements tirés de l'évaluation d'impact
• Poursuivre la surveillance et les évaluations de la performance

Source : d'après la Banque mondiale (2007b, p. 18).

Phase 1. *Le projet pilote d'une intervention novatrice et encore jamais mise en œuvre pour renforcer les moyens de subsistance des jeunes est sur le point d'être lancé. Quel type d'évaluation est le plus approprié ?*

Au tout début de la mise en œuvre d'un programme, nous devons généralement nous assurer que tout se déroule comme prévu. À ce stade, il n'est pas recommandé de conduire une évaluation d'impact car les résultats ne reflèteraient pas la qualité réelle du programme. Il est plus approprié de se concentrer sur le suivi et l'évaluation du processus jusqu'à ce que le programme soit pleinement réalisé et que les problèmes de mise en œuvre courants survenant lors de la mise en place d'un nouveau programme soient résolus. Les méthodes de collecte de données qualitatives (par exemple, entretiens avec des informateurs clés, groupes de discussion) peuvent être particulièrement utiles durant ces premières phases, car elles peuvent apporter des réponses à la question de savoir pourquoi certains éléments marchent ou ne marchent pas comme prévu ? On appelle souvent cette phase pilote initiale du projet « étude de faisabilité », destinée à obtenir une « validation du concept » ; c'est-à-dire, à indiquer si le projet peut effectivement être mis en œuvre comme prévu.

Phase 2. *L'intervention est menée depuis une année, et les problèmes liés au lancement de la mise en œuvre ont été résolus. Le suivi montre que les bénéficiaires sont satisfaits du programme. Devrions-nous étendre le programme ou le reproduire ailleurs ?*

Le moment est peut-être venu d'engager une évaluation d'impact. Le programme est en cours, et nous sommes assurés de la qualité de la mise en œuvre. Une évaluation d'impact nous permettra de confirmer que le programme a un effet sur les résultats escomptés. Nous pouvons également utiliser l'évaluation d'impact pour comparer l'efficacité des programmes déclinés autrement (par exemple, combinaison différente d'activités, intensité différente des activités) si nous ne sommes toujours pas sûrs de certains éléments de la conception. L'évaluation nous aidera aussi à comprendre certains effets potentiels qui n'étaient pas prévus (positifs ou négatifs). À partir des informations obtenues au moyen de l'évaluation d'impact, nous pouvons décider si des fonds importants doivent être investis ou non dans le programme.

Phase 3. *L'évaluation d'impact a révélé des résultats très positifs dans l'ensemble. Avons-nous encore besoin de conduire des évaluations ?*

Si les résultats positifs ne signifient pas que le programme serait aussi efficace dans des contextes différents, nous pouvons maintenant être relativement sûrs de la pertinence de notre théorie sur le changement et de la combinaison des activités. Nous disposons d'une bonne base pour étendre le programme à plus de participants ou pour le reproduire dans des contextes analogues. À moins que nous ne voulions modifier de manière importante notre intervention, il ne sera probablement pas nécessaire de conduire une autre évaluation d'impact. Nous devons toutefois être certains de maintenir à un niveau élevé la qualité de la mise en œuvre et de remplir nos objectifs. Le suivi à tous les stades, y compris au stade des résultats, doit demeurer un élément fondamental de notre programme. En outre, menées à intervalles réguliers, les évaluations indépendantes de la performance peuvent nous aider à contrôler la pertinence et la qualité du programme au fil du temps.

Ressources

Il se peut que certaines méthodes d'évaluation, intéressantes par ailleurs, ne soient pas applicables si nous ne disposons pas des ressources humaines et financières nécessaires. Il est important d'évaluer les compétences et les fonds disponibles dans le cadre de notre projet ou dans notre organisation afin de nous assurer qu'ils répondent aux besoins de l'évaluation que nous envisageons de mener.

Compétences

Pour conduire des évaluations de qualité, il faut disposer de compétences spéciales qui font parfois défaut dans un programme ou une organisation. Dans ce cas, et pour garantir la neutralité, il est souvent utile d'employer des évaluateurs externes. Le tableau 4.3. récapitule les principales compétences requises pour conduire les différents types d'évaluations.

TABLEAU 4.3 Compétences requises pour chaque type d'évaluation

Compétences	Description	Performance	Processus	Impact	Analyse coûts-bénéfices
Conception de programme et suivi	• Bonne connaissance des programmes visant à renforcer les moyens de subsistance des jeunes • Expérience dans la conception de programmes • Connaissances générales en matière de techniques de collecte de données quantitatives et qualitatives • Bonne connaissance du pays • Diplôme universitaire en sciences sociales	!	!	✓	s. o.
Collecte de données quantitatives	• Formation spécialisée en conception et réalisation d'enquêtes • Connaissances en analyse de données quantitatives • Compétences dans la gestion de projets pour la constitution et la direction d'une équipe de recenseurs • Diplôme universitaire en sciences sociales	✓	✓	!	s. o.
Analyse de données quantitatives	• Formation spécialisée en statistiques ou économétrie • Master ou doctorat en économie, santé publique ou dans un domaine connexe	✓	✓	!	✓
Collecte de données qualitatives	• Formation spécialisée dans la mise en œuvre de techniques qualitatives • Master ou doctorat en sociologie, anthropologie ou psychologie	✓	!	✓	s. o.
Analyse de données qualitatives	• Formation spécialisée en codage et analyse de données qualitatives • Master ou doctorat en sociologie, anthropologie ou psychologie	✓	!	✓	s. o.
Estimation	• Formation spécialisée en évaluation des coûts et des bénéfices des projets de services aux personnes • Master ou doctorat en économie, santé publique ou dans un domaine connexe	✓	s. o.	✓	!

! *requis ;* **✓** *avantage*

Financement

Les différences en termes d'envergure et de forme de la collecte et de l'analyse de données expliquent la grande variété des coûts d'évaluation. Se fonder sur la recherche documentaire et les entretiens réalisés avec des informateurs clés est naturellement beaucoup moins onéreux que concevoir de nouvelles enquêtes et les conduire auprès d'un grand nombre de personnes. Les évaluations de la performance représentent donc généralement le type d'évaluation le moins cher, alors que les évaluations d'impact tendent à être les plus coûteuses (voir tableau 4.4.).

TABLEAU 4.4 Estimation des coûts pour les différents types d'évaluation

Type d'évaluation	Coût	Facteurs influençant les coûts
Évaluation de la performance	USD 10 000–30 000	Ampleur de l'évaluation et salaire de l'évaluateur
Évaluation du processus	USD 10 000–60 000	Mêmes facteurs que pour l'évaluation de la performance, mais souvent utilisation de plus d'outils de collecte de données ; l'évaluation peut donc prendre plus de temps.
Évaluation d'impact	USD 15 000–1 million+	Les coûts varient considérablement en fonction de la méthode utilisée : plus de données sont collectées, plus l'évaluation est coûteuse (voir notes 6 et 7 pour plus de détails).
Analyses coûts-efficacité et coûts-bénéfices	USD 10 000–30 000	Les coûts varient selon que les bénéfices ont été précédemment estimés ou non et que les données sont facilement disponibles ou non.

Lorsque toutes les données sont facilement disponibles, les évaluations d'impact peuvent ne coûter que 15 000 dollars des États-Unis (USD), bien que dans la plupart des cas le coût dépassera les 100 000 dollars. Les évaluations d'impact peuvent sembler irréalisables pour des programmes dotés de budgets modestes. Mais leur coût peut se justifier si l'intervention est menée sur une longue période ou sur une vaste échelle. En outre, l'organisation d'exécution ne doit pas toujours prendre en charge le coût total d'une évaluation d'impact ; elle peut solliciter une aide financière pour conduire les évaluations (voir note 7 pour plus de détails sur le financement d'une évaluation d'impact).

Le contexte politique

Différents acteurs au sein et en dehors de notre organisation peuvent avoir des intérêts rivaux quant à la nécessité ou non de conduire une évaluation, quant aux questions à examiner, au type et à la méthode d'évaluation, à la stratégie de collecte de données et à la personne, le cas échéant, qui devrait être employée pour entreprendre l'évaluation. Tous ces facteurs peuvent se traduire par des pressions exercées pour le choix d'une évaluation et influencer la pertinence et la qualité de l'étude prévue. Les pressions peuvent aller d'une allusion subtile au fait que certaines questions ne devraient pas être examinées à une désapprobation catégorique de la part des pouvoirs publics quant au fait d'interroger certains groupes de jeunes, certaines familles ou communautés.

Il est donc important d'essayer de comprendre les divers intérêts et la situation politique qui caractérisent le contexte dans lequel l'intervention aura lieu. Les questions indiquées ci-après nous aideront à entreprendre notre analyse :

- Quel est le contexte politique local, et comment les pouvoirs sont-ils répartis ?
- Quelles sont les relations entre les bénéficiaires, les responsables de programme, les décideurs, les donateurs et autres acteurs ?
- Quels sont les intérêts et les motivations de chaque groupe d'acteurs pour influencer la conduite de l'évaluation et la conception du programme ? Par exemple, si le programme cible uniquement un groupe particulier de jeunes, ceux qui sont exclus auront un motif pour influencer le programme et l'évaluation de façon à profiter eux aussi des avantages du programme.
- Si l'évaluation montre que le programme a des effets, qui sont les potentiels gagnants et perdants de toute réforme des programmes ou des politiques qui pourrait découler de l'évaluation ?
- L'environnement local permettra-t-il la conduite d'une évaluation rigoureuse et indépendante, et aidera-t-il les évaluateurs à publier leurs résultats fondés sur des faits, quelles qu'en soient les conséquences sur le plan politique ?

Une ONG internationale et son partenaire local au Brésil ont décidé de conduire une évaluation d'impact sur un programme de formation-aptitude à l'emploi en faveur des jeunes, qu'ils mettaient en œuvre conjointement. Après quelques désaccords, les conditions d'admissibilité au programme ont finalement été fixées, et le choix de la sélection aléatoire des participants a notamment été fait. Cependant, le partenaire local avait conclu précédemment un accord avec une entreprise privée qui voulait influencer les décisions portant sur le choix des jeunes qui participeraient au programme, ce qui influencerait toute évaluation. Ce conflit a empêché la conduite effective de l'étude.

Essayer de comprendre les préoccupations des acteurs en entretenant un dialogue permanent et ouvert peut nous aider à trouver des moyens de faire face aux pressions et aux intérêts rivaux, et à rallier des soutiens pour mener l'évaluation. En outre, il est généralement utile de faire appel à des évaluateurs extérieurs qui, en plus d'apporter un ensemble de compétences spécifiques, peuvent plus facilement maintenir leur indépendance.

Types de programmes qui justifient habituellement la conduite d'une évaluation d'impact

Si les évaluations de la performance et du processus, ainsi que les analyses coûts-efficacité, font partie de tout programme, les évaluations d'impact et les analyses coûts-bénéfices devraient être entreprises de manière plus sélective. Selon Gertler et al. (2011), l'effort supplémentaire et les ressources nécessaires pour conduire des évaluations d'impact sont plus faciles à mobiliser lorsque le projet est : 1) stratégiquement pertinent et de nature à influencer les choses ; 2) novateur ou encore jamais réalisé ; et 3) reproductible.

Stratégiquement pertinent et de nature à influencer les choses

À quel point les résultats doivent-ils être importants pour servir de base aux programmes et aux politiques futures, ou au dialogue politique ? Si les enjeux d'une intervention sont importants (par exemple, parce qu'un programme nécessite des ressources considérables et qu'il couvre, ou pourrait être élargi pour couvrir, un grand nombre de personnes), alors il faudrait envisager une évaluation d'impact. Il en va de même pour les nouvelles initiatives ou les programmes existants lorsque nous devons décider de les poursuivre, de les étendre ou d'y mettre fin. En fait, même une évaluation d'impact coûteuse peut être très rentable, puisque les résultats qui en découlent peuvent contribuer à améliorer notablement la performance des programmes. Dans le cas d'initiatives de grande ampleur, même des améliorations mineures peuvent permettre à l'organisation d'exécution de faire des économies importantes (Banque mondiale 2009).

Novateur ou encore jamais réalisé

Quel est l'état actuel des preuves ou des connaissances acquises sur les effets du programme proposé ? Si l'on en sait peu sur l'efficacité du type d'intervention, à l'échelle mondiale ou dans un contexte particulier, une évaluation d'impact peut apporter des connaissances précieuses à notre organisation et aux acteurs concernés. C'est le cas de la plupart des programmes visant à renforcer les moyens de subsistance des jeunes, pour lesquels les preuves restent minces. Lorsqu'il y a peu de preuves disponibles, si toutefois il y en a, il est généralement préférable de commencer par un projet pilote qui comprend une évaluation d'impact. Même si l'on dispose de preuves concernant un certain type d'intervention, une évaluation d'impact peut toujours être justifiée si le programme est mis en œuvre dans un contexte différent ou qu'il inclut des aspects novateurs qui n'ont pas encore été testés.

[Ressources en ligne]

Manque de données et questions sur lesquelles pourrait porter la recherche en vue de l'évaluation d'impact.

http://www.iyfnet.org/gpye-m&e-resource3

ENCADRÉ 4.3 Le manque de preuves quant aux programmes axés sur les moyens de subsistance des jeunes

Les principes généraux indiqués ci-après doivent être interprétés avec prudence, nous pensons toutefois que les preuves existantes sur les programmes visant à renforcer les moyens de subsistance des jeunes semblent particulièrement fragiles dans ces domaines :

Types de programmes : La plupart des évaluations qui existent portent sur la formation et l'amélioration des compétences, et les preuves sont relativement rares sur tous les autres types d'interventions, tels que les emplois subventionnés en faveur des jeunes, les services de recrutement, l'entrepreneuriat des jeunes, les services financiers visant également les jeunes et les programmes ciblant les groupes exclus.

Éléments de la conception : On en sait peu sur la relative efficacité des programmes déclinés autrement. Pour chaque type de programme, quel est l'effet lorsque l'on adopte des éléments différents, des pédagogies différentes, des dosages et des moyens d'exécution différents ?

Contexte : Les preuves sur les programmes visant à renforcer les moyens de subsistance des jeunes sont particulièrement rares au Moyen-Orient et en Afrique du Nord, en Asie et en Afrique subsaharienne. En outre, on a besoin de plus de preuves pour savoir quelles interventions et quels éléments de la conception sont les mieux adaptés au contexte rural *versus* urbain, aux cadres informels *versus* formel, ou dans les situations d'après-conflit et les États fragiles.

Bénéficiaires : Dans quelle mesure différents types de programmes touchent-ils les jeunes différemment selon la tranche d'âge, le sexe, le niveau d'éducation et le statut socioéconomique ? Quel est le type de programme le plus efficace avec les groupes défavorisés ? Et quels sont les effets de propagation positifs ou négatifs de ces interventions sur les pairs, les familles et les communautés ?

Résultats : Quels sont les effets de ces programmes non seulement sur l'emploi et le marché du travail, mais également sur les comportements à risques, l'engagement civique, la constitution d'une famille, la santé mentale, etc. ? En outre, il n'existe pratiquement pas de preuves sur les effets à long terme de la plupart des interventions.

Pour un examen des preuves existantes, voir le « Youth Employment Inventory » (inventaire sur l'emploi des jeunes) (www.youth-employment-inventory.org) et Cunningham, Sanchez-Puerta et Wuermli (2010).

Reproductible

Dans quelle mesure et dans quelles circonstances un programme pilote ou de petite envergure donnant des résultats positifs pourrait-il être étendu ou reproduit auprès de groupes de population différents ? Si un projet d'intervention est extrêmement spécifique et cible un contexte particulier et restreint, alors une évaluation du processus qui contribuerait à une mise en œuvre sans heurts du programme serait probablement suffisante. Si toutefois le programme peut être étendu ou exécuté dans des contextes différents, alors une évaluation d'impact est une mesure importante, qui peut permettre de justifier qu'un programme soit reproduit.

Le tableau 4.5. présente un récapitulatif des principaux types d'évaluations.

TABLEAU 4.5 Aperçu des principaux types d'évaluation

	Évaluation de la performance	Évaluation du processus	Évaluation d'impact	Analyses coûts-efficacité et coûts-bénéfices
Quelles sont les principales questions auxquelles répond ce type d'évaluation ?	• Les objectifs du programme sont-ils clairement définis ? • La conception du programme est-elle appropriée pour réaliser les objectifs ? • Dans quelle mesure les objectifs du programme ont-ils été atteints ? • Faut-il changer les priorités ?	• Les ressources et les systèmes (gestion, information, etc.) dont on dispose sont-ils adéquats ? • Le programme est-il mis en œuvre conformément au projet prévu ? • Quelles sont les mesures effectivement prises et les activités réalisées pour fournir un produit ou un service ? • Qu'est-ce que les bénéficiaires ou autres acteurs savent ou pensent du programme ?	• Dans quelle mesure la situation des participants a-t-elle changé à la suite de l'intervention ? • Y a-t-il des conséquences imprévues, positives ou négatives, pour les participants au programme?	• Les coûts du programme sont-ils justifiés par rapport à des interventions semblables ? • Les coûts totaux du programme sont-ils justifiés en termes d'avantages obtenus ?
Quand cette évaluation peut-elle être conduite ?	Elle peut être conduite durant les premières phases de la mise en œuvre, pour les examens à mi-parcours ou au terme du projet.	Elle peut être conduite à tout moment, une fois ou à intervalles réguliers, pour vérifier que la mise en œuvre est en bonne voie ou pour comprendre des problèmes particuliers concernant le déroulement du programme.	Elle devrait être conçue lors de la planification du projet, mais les résultats finals ne seront généralement pas disponibles avant la fin du projet (ou avant la phase suivante).	Elle est généralement conduite au cours d'une analyse ex-ante pour déterminer si le programme vaut la peine d'être mis en œuvre ou poursuivi, ou au terme du programme pour déterminer les coûts finals.
Combien de temps l'évaluation prend-elle ?	1 à 3 mois (davantage si elle comprend l'analyse ex-ante/ex-post).	1 à 6 mois	• 6 mois au minimum (évaluation rétrospective) • 12 à 24 mois (évaluation prospective)	1 à 3 mois
Quelles collectes de données et analyses des données sont nécessaires ?	Étude préliminaire des documents existants et visites de terrain effectuées de manière sélective, complétées éventuellement par l'analyse des données de surveillance, les entretiens avec les bénéficiaires et les acteurs concernés, les mini-enquêtes, les groupes cibles, etc.	Mélange d'entretiens avec le personnel affecté à la mise en œuvre du programme et les bénéficiaires, d'enquêtes de satisfaction auprès des utilisateurs, d'examens des dossiers, d'observations directes, de groupes cibles et d'analyse des données de surveillance.	Analyse statistique et économétrique des données d'enquête et des données administratives, de préférence combinée avec l'analyse de données qualitatives.	Étude préliminaire des documents existants sur le programme et de la littérature pertinente, et entretiens avec des informateurs clés.
Qui conduit l'évaluation ?	Habituellement un évaluateur indépendant (mais qui peut aussi être interne).	Évaluateur interne ou indépendant	Équipe d'évaluation indépendante, comprenant par exemple un responsable d'évaluation, un coordonnateur sur le terrain, un cabinet d'études.	Évaluateur indépendant (peut être le même que pour l'évaluation de la performance ou l'évaluation d'impact).
Quelles compétences sont nécessaires ?	Analyse de programmes ; méthodes quantitatives simples si possible	Analyse des procédures, méthodes quantitatives et qualitatives	Analyse statistique et économétrique ; méthodes qualitatives si possible	Estimation et analyse économique des coûts et des bénéfices du programme
Quels sont les coûts ?	De 10 000 à 30 000 USD	De 10 000 à 60 000 USD	Les coûts peuvent aller de 15 000 à 1 million de dollars ou plus, selon l'envergure et la complexité du programme.	De 10 000 à 30 000 USD
Quels programmes sont les plus appropriés pour cette évaluation ?	Tout programme	Tout programme	Les programmes : • novateurs ou encore jamais réalisés • stratégiquement pertinents et de nature à influencer les choses • reproductibles	• Coûts-efficacité : tout programme • Coûts-bénéfices : mêmes programmes que pour l'évaluation d'impact

Source : d'après Rubio (2011).

Points essentiels

1. Le point de départ de toute évaluation est la définition de nos objectifs d'apprentissage. Il faut pour cela formuler des questions d'évaluation à tous les niveaux de la chaîne de résultats et classer par ordre de priorité les questions les plus pertinentes. En général, les questions d'évaluation sont *descriptives, normatives* ou portent *sur les causes et les effets.*
2. Le choix de la stratégie d'évaluation dépend des questions d'évaluation. Les besoins en informations purement descriptives ne nécessitent pas forcément la conduite d'une évaluation ; le suivi peut suffire. Les questions normatives trouvent le plus souvent une réponse dans les *évaluations du processus* ou *de la performance.* Si les questions portant sur les causes et les effets sont la priorité, alors les *évaluations d'impact* sont nécessaires. Les *analyses coûts-efficacité* et *coûts-bénéfices* permettent de savoir si les coûts engendrés par une intervention sont justifiés.
3. Seules les évaluations d'impact – celles qui permettent d'établir un contrefactuel valable – nous permettent de *prouver* si un programme a donné de bons résultats et génèrent des informations qui peuvent être généralisées au-delà de l'intervention elle-même. Cela les différencie de tous les autres types d'évaluations et fait d'elles un instrument essentiel pour évaluer les interventions destinées à renforcer les moyens de subsistance des jeunes.
4. Le choix d'un type d'évaluation approprié dépend du contexte dans lequel le programme se déroule. Il est donc crucial de comprendre si les coûts en termes d'argent, de personnel et de temps pour chaque évaluation sont appropriés pour une intervention donnée.
5. Étant donné que les évaluations d'impact tendent à nécessiter le plus de ressources, en comparaison avec les autres types d'évaluations, il faudrait les conduire de manière sélective afin de répondre à des questions d'ordre stratégique ou d'évaluer les interventions pilotes novatrices lancées pour tester une nouvelle approche prometteuse.

Étude de cas du NUSAF : Décider de conduire ou non une évaluation d'impact

Questions d'évaluation

Le principal objectif d'apprentissage pour le NUSAF était d'estimer l'effet de causalité de la participation aux programmes de formation professionnelle sur les moyens de subsistance économiques et l'insertion sociale. Les questions intéressant le NUSAF étaient de savoir si le programme « Youth Opportunities Program », visant à améliorer les perspectives d'emploi des jeunes, contribuait à :

- accroître le nombre de créations d'entreprise ;
- abaisser les niveaux de chômage ;
- augmenter le nombre d'heures travaillées contre rémunération ;
- améliorer l'insertion au sein de la communauté et à apaiser les conflits ;
- réduire la pauvreté ;
- accroître le bien-être psychosocial.

Étant donné la nature de ces questions portant sur les causes et les effets, l'évaluation d'impact était la méthode d'évaluation de choix.

Le NUSAF voulait également évaluer les effets du programme sur les organismes de formation locaux. Vu que les évaluations d'impact ne permettent pas de procéder facilement à une telle évaluation, il a été décidé que cette estimation entrerait dans le cadre de la surveillance du programme.

Contexte dans lequel se déroule le NUSAF

Le NUSAF étant un projet financé par la Banque mondiale et solidement soutenu par le Gouvernement ougandais en ce qui concerne l'évaluation d'impact, le contexte était favorable à la conduite d'une telle évaluation.

- Moment : la stratégie d'évaluation a été prévue dès le début du programme, ce qui a permis d'avoir la flexibilité nécessaire pour planifier une évaluation d'impact rigoureuse.
- Ressources : les ressources nécessaires ont pu être affectées à l'avance, et une équipe externe qualifiée a pu être recrutée pour conduire l'évaluation.
- Contexte politique : faire de l'évaluation une priorité dès le début a favorisé le dialogue avec les acteurs concernés et motivé leur soutien.

Caractéristiques du NUSAF qui justifieraient une évaluation d'impact

Le programme « Youth Opportunities Program » était un vaste programme de dons en espèces conçu et mis en œuvre par le Gouvernement ougandais. Vu l'ampleur et l'influence du programme, ainsi que l'espoir de voir le programme reconduit à l'avenir, on a estimé que l'évaluation du programme était une excellente manière de renforcer les connaissances locales et mondiales sur les programmes de formation en échange de dons en espèces. Bien que ces types de programmes soient de plus en plus mis en œuvre, ils n'ont généralement pas encore été évalués. En outre, le fait que le programme soit exécuté par le Gouvernement suggère qu'un tel programme est modulable et reproductible dans d'autres pays.

Source : d'après Blattman, Martinez et Fiala (2009).

À lire

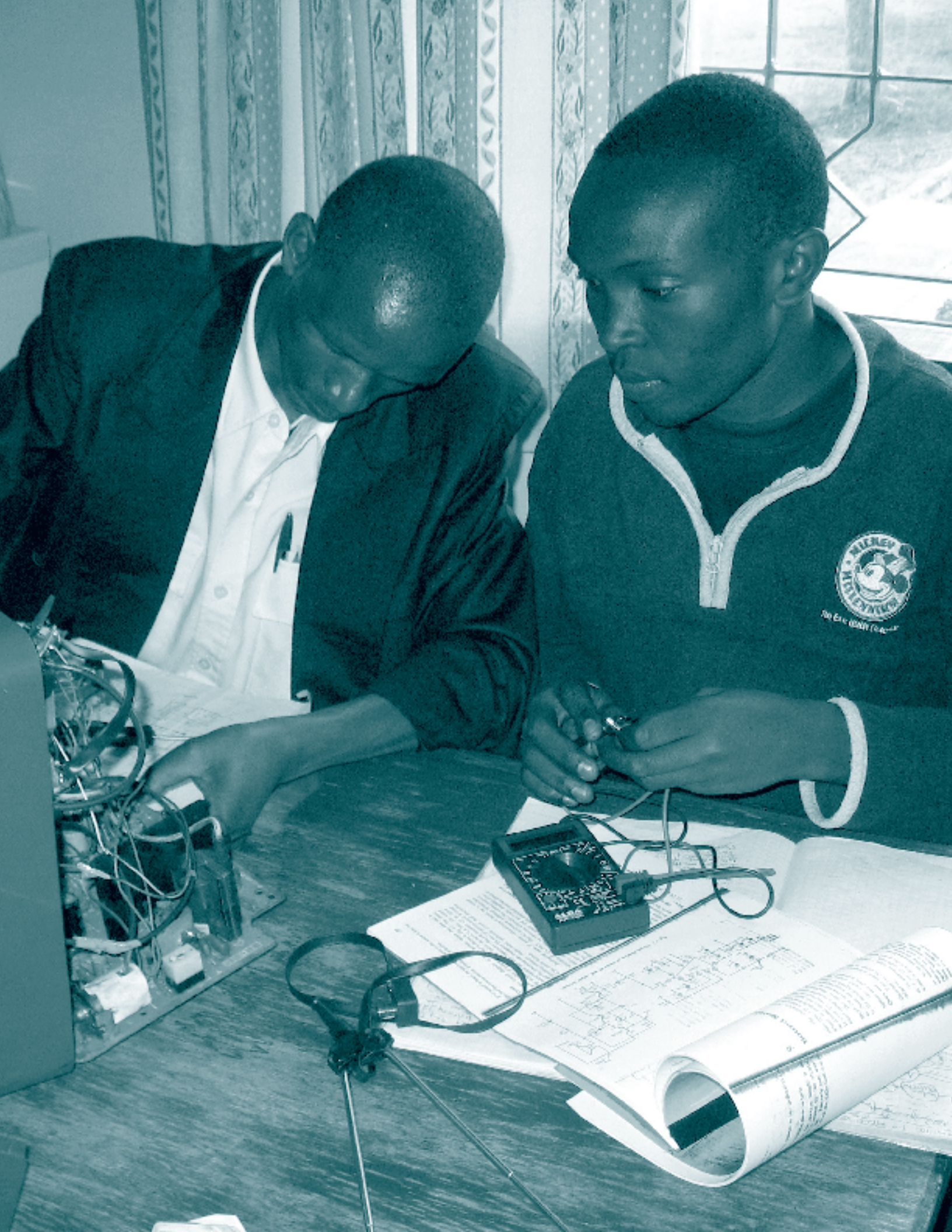
Imas, L., et Rist, R., 2009, *The Road to Results: Designing and Conducting Effective Development Evaluations*, Washington, D.C., La Banque mondiale. Les chapitres 6 et 7 sont en rapport avec cette note.

Rubio, G., 2011, "The Design and Implementation of a Menu of Evaluations." PREM Notes, The Nuts and Bolts of M&E Systems, No. 6, Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTPOVERTY/Resources/335642-1276521901256/premnoteME6.pdf

Notes

PARTIE II
RENFORCER LES CONNAISSANCES SUR LE PROGRAMME AU MOYEN DE L'ÉVALUATION D'IMPACT

Note : La IIe partie de ce guide est fondée sur l'expérience du terrain acquise par l'un des auteurs ainsi que sur des matériaux élaborés à l'origine par Duflo, Glennerster et Kremer (2006) ; Khandker, Koowal et Samad (2010) ; et Gertler et al. *(2011). Certains de ces matériaux et illustrations ont été adaptés au domaine des moyens de subsistance des jeunes. Cette IIe partie offre une présentation plus concise des méthodes d'évaluation d'impact. Les auteurs reconnaissent les contributions apportées par les spécialistes susmentionnés et recommandent la lecture de la publication intitulée* Impact Evaluation in Practice *(Gertler* et al. *2011) aux personnes qui sont intéressées par une description détaillée du contenu proposé dans cette IIe partie.*



NOTE 5 : Prouver l'impact d'un programme

Le scepticisme rigoureux est une force créatrice,
car la plupart du temps les dégâts sont causés par
des personnes présomptueuses qui pensaient connaître
la réponse, alors qu'elles n'en savaient rien.

— William Easterly

Les bonnes intentions ne suffisent pas. Nous avons besoin de savoir que nous améliorons effectivement la vie des personnes et que nous ne faisons pas plus de mal que de bien sans même le savoir. La preuve nous en est donnée par les évaluations d'impact, qui, à la différence des autres types d'évaluation, apportent des éléments de preuve scientifiques sur l'efficacité d'un programme.

Dans cette note, nous étudierons la question fondamentale de l'évaluation d'impact : « Comment pouvons-nous être sûrs que les changements observés dans les résultats ont été produits par notre intervention ? ». Nous montrerons que pour mesurer l'impact il faut estimer ce qui se serait passé en absence d'intervention. On peut procéder à ces estimations par des techniques d'évaluation expérimentale ou quasi expérimentale qui reposent sur l'analyse d'un groupe témoin. Nous montrerons également pourquoi les deux techniques les plus courantes – la comparaison des participants avant et après l'intervention et la comparaison des participants avec des non-participants sélectionnés subjectivement – ne peuvent pas fournir d'estimations fiables sur l'efficacité d'un programme.

La difficulté de déterminer la cause des résultats observés

Les évaluations d'impact nous aident à répondre à des questions très spécifiques concernant notre programme. Comme nous l'avons vu dans la note 4, elles visent à indiquer si une intervention (la cause) améliore les résultats parmi les bénéficiaires (l'effet). Par exemple :

- Notre programme de formation professionnelle permet-il d'accroître les revenus des participants ?
- Notre programme d'enseignement en milieu scolaire consacré à l'entrepreneuriat contribue-t-il à accroître les taux de scolarisation jusqu'à la fin des études secondaires et à motiver les élèves à poursuivre des études supérieures ?
- Notre programme d'aide à la création d'entreprises individuelles favorise-t-il la création d'entreprises et leur viabilité ?

Il peut être compliqué d'établir le lien de cause à effet entre les activités menées dans le cadre de notre intervention et les résultats que nous observons, car d'autres facteurs peuvent également influencer les résultats d'intérêt. Par exemple, le fait d'observer simplement que des entreprises ont été créées après la mise en œuvre de notre programme d'entrepreneuriat ne prouve pas l'efficacité de notre programme, car d'autres facteurs, tels que la situation économique locale ou la réglementation concernant la création d'entreprises, peuvent avoir changé et favorisé la création d'entreprises. De la même façon, l'observation d'un déclin dans la création d'entreprises après notre intervention ne signifie pas nécessairement que c'est notre intervention qui a *causé* ce déclin ; il peut s'agir au contraire d'une détérioration d'autres conditions externes.

Le but des évaluations d'impact est précisément de surmonter cette difficulté qu'il y a à déterminer la cause des résultats observés en évaluant dans quelle mesure un projet particulier, *et uniquement ce projet*, a contribué à changer les résultats d'intérêt.

Qu'est-ce que l'« impact » exactement ?

[Définition]

Résultats obtenus avec le programme

– Résultats observés en absence de programme

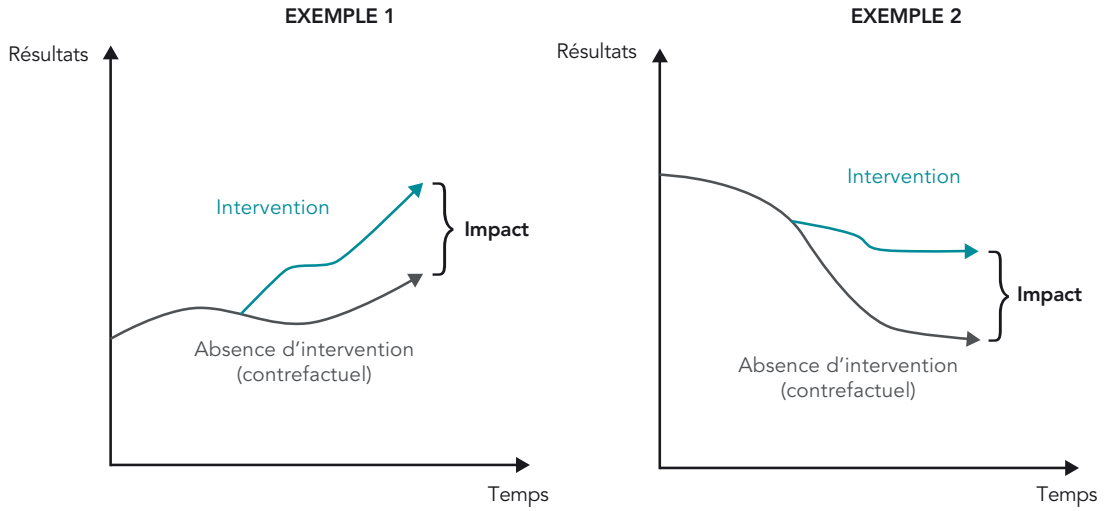
= **Impact**

Premièrement, nous devons éclaircir ce que nous entendons par *impact*. Le terme renvoie souvent aux objectifs du programme au plus haut niveau ou aux résultats relatifs aux changements survenus dans les conditions de vie en général, tels que la réduction de la pauvreté ou l'accroissement du bien-être des personnes et des ménages. Dans le cadre des évaluations d'impact, cependant, l'*impact* a une signification plus étroite et désigne le changement produit au niveau des résultats qui peut être directement attribué à notre programme. L'accent ici est mis sur « directement attribué », qui signifie que nous voulons savoir si les changements survenus dans les résultats que nous observons sont véritablement dus à notre intervention et à rien d'autre.

En clair, comme le montre la figure 5.1., l'impact d'une intervention est la différence entre les éléments suivants :

- les résultats observés dans le cadre de l'intervention ; et
- les résultats observés auprès de la même personne, des mêmes ménages, de la même communauté ou autre unité d'observation, en absence d'intervention. Ces résultats sont appelés le *contrefactuel* et renvoient à *ce qui serait arrivé au bénéficiaire si le programme n'avait pas été mis en œuvre*.

FIGURE 5.1 Illustration visuelle de l'impact d'un programme



Pour des raisons évidentes, il est impossible d'observer la même personne (ménage, école, etc.) après l'intervention et en absence d'intervention. S'il est possible d'observer les résultats obtenus auprès de ces jeunes qui participent à notre programme, il est impossible de savoir dans quelle situation ils auraient été en absence de programme. Ce qui fait que nous ne pouvons pas savoir avec certitude ce qui leur serait arrivé s'ils n'avaient pas participé à notre programme. Comme nous ne pourrons jamais connaître les véritables éléments contrefactuels, une estimation doit suffire.

Comment pouvons-nous estimer le contrefactuel ??

Pour estimer les éléments contrefactuels, nous identifions des *groupes témoins*, également appelés *groupes de contrôle*. Le groupe des participants au programme est connu sous le nom de *groupe de traitement*. Un groupe témoin de qualité possède les mêmes caractéristiques que le groupe de traitement, excepté le fait que les membres du groupe témoin ne sont pas au bénéfice du programme.

Selon Gertler *et al.* 2011, le groupe de traitement et le groupe témoin devraient posséder les mêmes caractéristiques sur trois plans au moins :

1. **Ils devraient être identiques en termes de caractéristiques observables et non observables.** Les caractéristiques observables font référence à l'âge, au sexe, au niveau d'éducation, au statut socioéconomique, aux caractéristiques familiales, à la situation au regard de l'emploi, etc. Les caractéristiques non observables comprennent la motivation, l'intérêt, les préférences, le degré de soutien familial et d'autres facteurs. Si les sujets du groupe de traitement ne doivent pas tous être identiques aux sujets du groupe témoin, les deux groupes devraient être en moyenne similaires.
2. **Le groupe de traitement et le groupe témoin devraient réagir au programme de la même façon.** Par exemple, la probabilité que les résultats, tels que les compétences ou le revenu, s'améliorent devrait être la même pour les sujets du groupe de traitement que pour ceux du groupe témoin.
3. **Le groupe de traitement et le groupe témoin devraient bénéficier sur le même pied d'égalité des autres interventions.** Par exemple, les deux groupes devraient avoir un accès égal aux autres services d'aide fournis par les autorités locales, les ONG, etc.

Lorsque les conditions énumérées ci-dessus sont égales entre les deux groupes, alors seule l'existence de l'intervention expliquera toute différence dans les résultats. Dans ce cas, l'effet de causalité

[Définition]

Un **groupe témoin** est un groupe qui possède les mêmes caractéristiques que le groupe des participants, excepté le fait que les sujets du groupe témoin ne sont pas au bénéfice du programme. Les termes « groupe témoin » et « groupe de contrôle » sont souvent utilisés indifféremment, bien que, à proprement parler, le second terme soit applicable uniquement dans le cadre d'évaluations expérimentales (voir plus bas). Nous utiliserons le terme générique *groupe témoin* dans le présent document.

[Définition]

Les **biais de sélection** apparaissent habituellement lorsque les participants au programme et les non-participants présentent des caractéristiques différentes qui ne peuvent pas être observées, mais qui influencent la décision de la personne de participer ou non au programme, ainsi que les résultats d'intérêt.

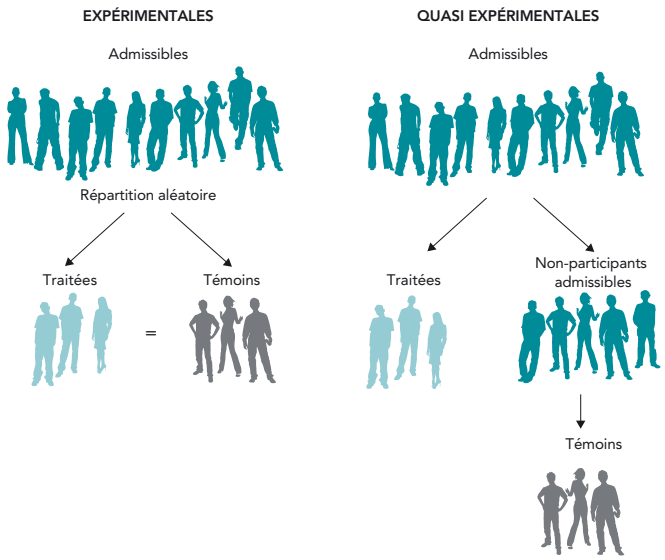
du programme peut être prouvé. Si en revanche le groupe témoin diffère de manière significative du groupe de traitement, nous sommes face à un *biais de sélection* (ou erreur systématique), qui invalidera nos mesures de l'impact. Ces biais de sélection renvoient au fait que des différences profondes entre le groupe de traitement et le groupe témoin expliquent la raison pour laquelle nous observons des résultats différents. Les biais de sélection apparaissent souvent lorsque le groupe témoin est composé de personnes qui ne remplissent pas les conditions requises pour bénéficier du programme (conditions fondées sur des caractéristiques observables) ou qui ont choisi de ne pas y participer (pour des raisons non observables).

Dans le cas de programmes de formation et de renforcement des moyens de subsistance, il est probable que les personnes qui se portent candidates soient différentes de celles qui ne sont pas candidates, et il n'est pas facile pour le chercheur d'observer ces différences. Par exemple, les candidats peuvent être plus motivés ou disposer de plus d'informations que les autres. Ces différences peuvent également signifier que les candidats, en moyenne, ont plus de chances sur le marché du travail que les autres, *indépendamment de la formation*. Dans ce cas, les résultats plus positifs parmi les bénéficiaires de la formation peuvent être dus à ces différences profondes et non pas à la formation que les personnes ont suivie dans le cadre du programme.

Techniques pour constituer des groupes témoins de qualité

Il existe en gros deux moyens de s'assurer que le groupe de traitement et le groupe témoin soient aussi semblables que possible : 1) avec les techniques expérimentales ; et 2) avec les techniques quasi expérimentales (voir figure 5.2.).

FIGURE 5.2 Techniques expérimentales et techniques quasi expérimentales



Techniques expérimentales

Dans les procédures d'évaluation expérimentales, la répartition des personnes dans les groupes se fait par *aléation*. Ainsi, si nous disposons d'un groupe de bénéficiaires potentiels (disons 500 jeunes, 500 écoles, etc.), nous sélectionnons par aléation certains d'entre eux (par exemple 250) qui bénéficieront du programme. Il s'agit du groupe de traitement. Les autres n'en bénéficieront pas ; c'est le groupe témoin. Si l'aléation est effectuée correctement, il est probable que les deux groupes soient très similaires du point de vue : 1) des caractéristiques observables et non observables ; 2) de la manière

dont ils réagiraient au programme ; et 3) de leur exposition à d'autres interventions. Les évaluations menées à l'aide de cette technique, ou autres variantes, sont communément appelées « essais contrôlés randomisés ». Voir l'encadré 5.1. pour des considérations d'ordre éthique sur l'aléation.

ENCADRÉ 5.1 L'aléation est-elle éthique ?

Certains concepteurs de programmes sont réticents à l'idée de répartir aléatoirement les potentiels bénéficiaires entre le groupe de traitement et le groupe témoin. Le souci généralement partagé est que l'évaluation conduit à refuser à des personnes dans le besoin des avantages apparemment évidents (tels que des possibilités de formation), ce qui ne serait pas éthique. En réalité, il est faux de supposer qu'une personne se verrait refuser un avantage si un programme n'a pas encore été correctement évalué. Dans les programmes qui n'ont pas été évalués, la répartition aléatoire peut en fait être plus éthique que d'autres méthodes de sélection pour les raisons suivantes :

- **Incertitudes quant à l'impact du programme.** Pour la plupart des programmes, il n'est pas clairement établi que le programme aura des effets positifs sur la personne et la communauté ou que les effets sont suffisamment importants pour justifier les coûts engendrés par le programme. Une intervention peut en fait avoir un effet nul ou mêmes des effets secondaires négatifs qu'on n'avait pas prévus. Par exemple, les programmes visant exclusivement les filles peuvent accroître la violence sexiste. Un programme de micro-finance destiné aux jeunes peut aggraver la situation économique des participants s'ils ne sont pas en mesure de rembourser leur emprunt. Même un programme de formation, s'il est mal conçu, peut en réalité réduire les perspectives d'emploi. Lorsque des effets positifs sont enregistrés (par exemple, augmentation du revenu des participants de 100 dollars), ce peut être à un prix très élevé (par exemple, dépense de 1 000 dollars par personne), ce qui laisse penser qu'il vaudrait bien mieux consacrer cet argent à autre chose. Ainsi, dans le cas d'interventions dont les effets et le rapport coûts-bénéfices n'ont pas été suffisamment prouvés, il est justifié d'évaluer le programme en utilisant des groupes de traitement et des groupes témoins.
- **Contraintes budgétaires.** Dans la réalité, en raison des ressources limitées, il est rarement possible d'aider toutes les personnes dans le besoin. En effet, la plupart des programmes n'apportent des avantages et des services qu'à un nombre limité de bénéficiaires, ce qui signifie que d'autres personnes sont exclues, que ce soit explicite ou non. Par exemple, si un programme de formation destiné aux jeunes offre un nombre limité de places, certains jeunes bénéficieront de la formation alors que d'autres non. De la même façon, si une intervention est menée dans un district particulier, les jeunes satisfaisant aux conditions requises mais vivant dans d'autres districts en seront exclus. La L'aléation permet aux responsables de programmes de sélectionner des candidats parmi les participants potentiels d'une manière qui est juste et qui donne les mêmes chances de participation à tous. Si l'aléation est effectuée ouvertement (par exemple, à la manière d'une loterie lors d'un événement public), elle renforce également la transparence de la procédure de sélection et peut réduire les craintes au sein de la population quant à une sélection fondée sur des préférences personnelles ou politiques.

Il est également important de noter que les évaluations randomisées ne signifient pas nécessairement refuser des services à quiconque. Des détails seront fournis dans la note 6 sur des techniques d'évaluation différentes.

Techniques quasi expérimentales

L'aléation n'est pas toujours faisable ni souhaitée (voir encadré 5.2.). Dans ces cas-là, on peut utiliser les techniques quasi expérimentales pour isoler l'effet de notre intervention. Bien que ces techniques soient généralement moins fiables que les méthodes expérimentales, elles tendent à simuler le contrefactuel en recensant des non-participants qui sont aussi semblables que possible aux sujets du groupe de traitement. Pour ce faire, les méthodes quasi expérimentales sont habituellement fondées sur des outils et des analyses statistiques. Parmi les méthodes les plus courantes, on trouve l'approche par discontinuité de la régression, la méthode de la double différence et l'appariement (voir note 6 pour un examen plus détaillé).

ENCADRÉ 5.2 Exemples de cas où l'aléation n'est pas possible

- La mise en œuvre du programme a déjà commencé ; les bénéficiaires ont déjà été sélectionnés.
- Les ressources disponibles sont suffisantes pour tous les membres de la population remplissant les conditions requises. Il peut alors être contraire à la déontologie de refuser d'offrir des avantages ou des services aux seules fins de l'étude.
- Nous ne pouvons pas sélectionner un groupe témoin ni exclure quiconque du programme. Par exemple, une campagne médiatique télévisée ou radiophonique destinée à améliorer les compétences financières touche potentiellement chaque foyer, et il est impossible de savoir qui est devant son poste et qui ne l'est pas.
- L'intervention cible un nombre limité de groupes ou de communautés ayant des caractéristiques uniques.
- Il y a une opposition politique quant au fait de faire bénéficier un groupe et pas un autre d'une intervention.

Lorsque les conditions sont réunies pour constituer un groupe témoin de qualité, on dit que l'évaluation d'impact a une validité interne (voir encadré 5.3.).

ENCADRÉ 5.3 Validités interne et externe

Idéalement, les évaluations d'impact doivent satisfaire à deux critères :

1. La *validité interne* ; ce qui signifie que nous pourrons montrer l'effet de causalité. Pour ce faire, nous tenons compte de toutes les différences possibles entre le groupe de traitement et le groupe témoin, et sommes en mesure d'attribuer clairement à l'intervention les changements observés dans les résultats. Pour garantir cela, nous utilisons des techniques expérimentales ou quasi expérimentales (voir l'examen détaillé dans la note 6).

2. La *validité externe* ; ce qui signifie que nous pourrons généraliser les conclusions de l'évaluation. En effet, nous pouvons nous attendre à obtenir les mêmes résultats si nous inscrivons des groupes différents ou plus nombreux au programme. Pour garantir cela, nous avons besoin d'appliquer une stratégie adéquate destinée à sélectionner l'échantillon de personnes avec lesquelles nous travaillons (voir note 7 pour une discussion à ce sujet).

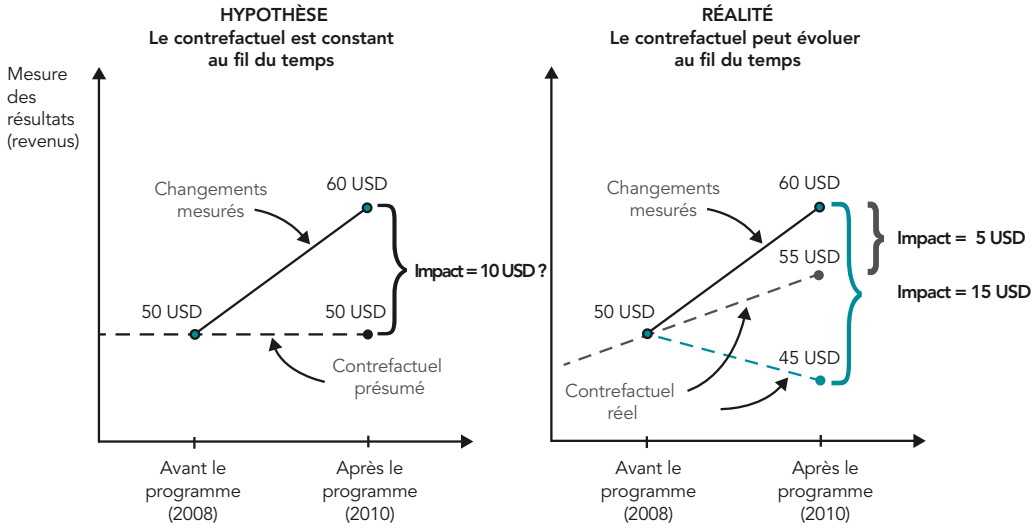
Contrefactuels contrefaits

Les deux techniques les plus couramment utilisées pour évaluer l'efficacité de nos programmes sont la comparaison des participants avant et après l'intervention, et la comparaison des participants avec les non-participants sélectionnés subjectivement. Ces techniques ne permettent pas de constituer un groupe témoin de qualité. Aussi ne peuvent-elles pas être considérées comme des méthodes d'évaluation d'impact valables, et les évaluations d'impact qui en résultent ne sont généralement pas crédibles. Voici pourquoi.

1er contrefactuel contrefait : comparaison des participants avant et après

Avec cette technique, nous utilisons les résultats obtenus avant l'intervention pour estimer le contrefactuel. Ainsi, nous supposons que si le programme n'avait jamais existé, les résultats obtenus par ces participants après le programme auraient été exactement les mêmes qu'avant le programme. Dans le cas d'un programme de formation, nous pouvons observer que le revenu mensuel des participants est passé de 50 dollars avant le programme à 60 dollars après le programme. Nous pouvons donc en conclure que l'impact du programme a été de 10 dollars par mois et par personne (voir figure 5.3., graphique de gauche).

FIGURE 5.3 Risques inhérents à la comparaison des résultats avant et après



Le Problème

L'hypothèse selon laquelle en absence de programme rien n'aurait changé n'est simplement pas fondée dans la plupart des cas. Beaucoup de choses peuvent se produire au cours de la période de mise en œuvre, en particulier lorsque les programmes s'étendent sur plusieurs années. Par exemple, les conditions économiques locales peuvent s'améliorer, ce qui favorise l'augmentation du nombre d'offres d'emploi et du niveau moyen des revenus ; les conditions climatiques favorables peuvent contribuer à augmenter les rendements et les revenus dans l'agriculture ; ou les autorités locales peuvent mettre en œuvre leur propre programme de travail contre revenus, ce qui accroît les revenus de nombreux jeunes. Si, en fait, l'environnement extérieur s'améliorait indépendamment du programme, alors les jeunes verraient leur revenuaugmenter de toute façon (disons 55 dollars par mois), et l'impact réel de notre intervention serait probablement bien moindre que celui estimé à l'aide d'une simple comparaison avant-après. Dans notre exemple, le gain serait de 5 dollars au lieu de 10 (voir figure 5.3., graphique de droite). Au contraire, si les conditions empiraient (disons que les jeunes gagneraient uniquement 45 dollars par mois en absence du programme), alors nous sous-estimerions l'impact véritable du programme en utilisant une comparaison avant-après.

Conclusion

Nombre de facteurs peuvent influencer dans le temps les effets des interventions visant à renforcer les moyens de subsistance des jeunes. Aussi l'évaluation des résultats d'un avant-programme est presque toujours une mauvaise estimation du contrefactuel. C'est pour cette raison que la comparaison avant-après n'est pas considérée comme une technique fiable pour démontrer l'impact d'un programme.

2e contrefactuel contrefait : comparaison des participants et des non-participants

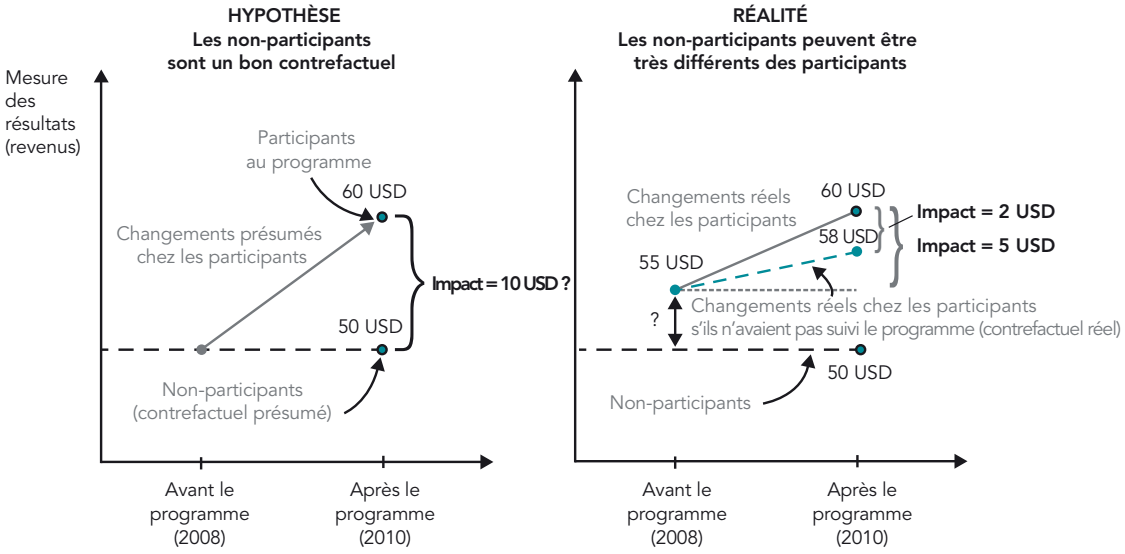
Avec cette technique, nous observons les résultats obtenus par les non-participants sélectionnés subjectivement au terme de l'intervention pour estimer le contrefactuel. Lorsque nous comparons les participants avec les non-participants, nous supposons que ces groupes sont très semblables par nature. Par exemple, les deux groupes possèdent les mêmes caractéristiques observables et non observables, réagiraient au programme de la même manière et sont exposés à d'autres interventions sur un même pied d'égalité.

[Conseil]

Les interventions ciblées à court terme visant à améliorer des attitudes, des connaissances et des compétences particulières (voir par exemple l'évaluation avant-après du programme de l'OIT *Know About Business* dans l'encadré 5.4. ci-après) sont un cas spécifique dans lequel les comparaisons avant-après peuvent fournir un contrefactuel relativement solide. Dans ce cas, les résultats peuvent parfois être d'une façon réaliste attribués à une intervention en particulier. Cependant, d'autres effets potentiels liés au programme, tels que des changements dans le comportement, l'emploi et le revenu, sont influencés par de nombreux facteurs, et une simple comparaison avant-après ne permet pas toujours de les estimer correctement.

Si l'on reprend notre exemple de programme de formation, nous mesurerions le niveau de revenu à la fois des participants et des non-participants à la fin de la formation. Supposons que nous trouvions que les participants gagnent 60 dollars par mois, alors que les non-participants n'en gagnent que 50, nous pourrions en conclure que l'impact de notre programme a été un gain de 10 dollars par mois et par personne (voir figure 5.4., graphique de gauche).

FIGURE 5.4 Risques inhérents à la comparaison des participants avec des non-participants



Le problème

Cette approche pose deux problèmes majeurs. Premièrement, l'hypothèse selon laquelle les deux groupes ont des résultats en moyenne équivalents au début du programme est peut-être erronée. Les participants peuvent avoir gagné plus ou moins avant le lancement du programme que les non-participants sélectionnés subjectivement. Si nous ne mesurons les résultats qu'à la fin du programme, il se peut que nous ne soyons pas en mesure d'en apprendre plus sur les conditions de départ. Les participants peuvent percevoir déjà au début du programme des revenus plus élevés que les non-participants (par exemple, 55 dollars) ; ainsi, le changement réel comparé avec notre observation au terme de la formation (60 dollars) serait de 5 dollars au lieu de 10 (voir figure 5.4., graphique de droite).

Deuxièmement, l'hypothèse selon laquelle les participants et les non-participants sont très similaires est généralement fausse. Il suffit de penser aux critères utilisés pour sélectionner les jeunes qui participeront au programme. Ce peut être par ordre d'arrivée. Dans ce cas, ceux qui disposent d'un meilleur accès à l'information concernant le programme, ceux qui habitent dans les environs, ceux qui sont encouragés par leurs parents, ou simplement ceux qui sont motivés pour y participer auraient plus de chances de suivre le programme. Autrement, les critères de sélection clairement définis, comme les résultats à un test de contrôle, les entretiens ou la qualité d'un projet de création d'entreprise, indiquent que nous voulons explicitement des participants qui soient différents des non-participants. Dans les deux cas, que nous le voulions ou non, les participants et les non-participants seront probablement différents, en moyenne, les uns des autres; il est par conséquent trompeur de comparer les deux groupes. En réalité, étant donné leur motivation potentiellement plus grande, leur meilleur accès à l'information, leur proximité avec les services, etc. – caractéristiques qui ne sont pas toujours facile à observer pour nous –, les jeunes qui ont participé à notre programme auraient

très bien pu améliorer leur situation même en absence d'intervention. Reprenons notre exemple : si les participants avaient gagné 58 dollars au bout d'un certain temps même sans avoir suivi notre programme, alors leurs revenus totaux au terme de la formation (60 dollars) refléteraient un impact de programme de 2 dollars seulement, et non pas de 10 dollars (voir figure 5.4., graphique de droite).

Conclusion

Il y a habituellement des raisons profondes à la participation ou non des personnes à un programme. Cela signifie que les participants et les non-participants sont fondamentalement différents les uns des autres, que nous puissions l'observer (par des résultats à des tests de contrôle) ou non (soutien familial, motivation). Ainsi, les non-participants sélectionnés subjectivement ne représentent presque jamais un bon contrefactuel permettant de comprendre comment les participants auraient fait en absence de programme. Par conséquent, on considère que la simple comparaison des participants et des non-participants sans recours à des techniques expérimentales ou quasi expérimentales n'est pas une bonne méthode pour prouver les effets d'un programme.

Bien que les contrefactuels contrefaits exposés plus haut ne soient peut-être pas utiles pour estimer l'impact – c'est-à-dire pour apporter une réponse aux questions sur les causes et les effets –, ils peuvent toujours servir dans nos programmes. En fait, il peut être intéressant pour la gestion du projet de recueillir au fil du temps des informations descriptives sur les participants et même les non-participants, car cela peut nous aider à mieux comprendre l'évolution de notre programme. Il est absolument légitime d'utiliser ces types de comparaison dans le cadre de notre surveillance ou de l'évaluation de la performance, pour autant que nous sachions ce que les résultats peuvent ou ne peuvent pas nous révéler (voir encadré 5.4. pour des exemples).

ENCADRÉ 5.4 Exemples choisis d'évaluations non expérimentales

Technique : comparaison avant-après

Programme de l'OIT *Know About Business*, Syrie
Assessing the Effect of Know About Business (KAB) on the Knowledge and Attitudes of Secondary School Students (2007)
http://www.syriatrust.org/site/images/files/KAB_Schools_Report_0708.pdf

Technique : comparaison des participants avec des non-participants

Junior Achievement, États-Unis d'Amérique
The impact on students of participation in JA Worldwide: Selected cumulative and longitudinal findings (2004)
http://www.ja.org/files/long_summary.pdf

Points essentiels

1. L'impact d'un programme est le changement observé dans les résultats qui peuvent être directement attribués à l'intervention. Pour comprendre l'impact, il nous faut isoler les effets du programme des autres facteurs influençant les résultats des bénéficiaires.
2. Pour mesurer l'impact d'un programme, il faut un contrefactuel, ce qui nous permet de savoir ce qui serait arrivé aux participants de notre programme en absence d'intervention.
3. Afin d'estimer ce qui serait arrivé aux bénéficiaires en absence de programme, nous constituons des groupes témoins qui possèdent le plus possible de caractéristiques semblables à celles des bénéficiaires. Si l'on peut constituer un groupe témoin de qualité, la comparaison des résultats obtenus par le groupe témoin et le groupe des bénéficiaires (groupe de traitement) reflète l'impact du programme.

4. Les techniques d'évaluation d'impact utilisées pour constituer des groupes témoins valables peuvent être de deux sortes. Avec les techniques expérimentales, on sépare aléatoirement la population remplissant les conditions requises en deux groupes : l'un qui bénéficiera du programme et l'autre qui n'en bénéficiera pas. Avec les techniques quasi expérimentales, on essaie de constituer un groupe témoin valable parmi les non-participants, qui reflète le groupe de traitement aussi fidèlement que possible.
5. Les simples comparaisons avant-après ainsi que la comparaison des participants avec des non-participants sélectionnés subjectivement ne fournissent pas d'estimations de l'impact qui soient crédibles. Les premières ne permettent pas de tenir compte des facteurs externes qui ont changé au fil du temps, la seconde ne permet pas de tenir compte des caractéristiques (souvent impossibles à observer) qui influencent le placement dans un programme. Ces deux types de comparaison peuvent toutefois être utiles pour recueillir des informations descriptives dans le cadre de notre système de surveillance..

Étude de cas du NUSAF : identifier un contrefactuel

L'identification du contrefactuel a été particulièrement difficile pour l'évaluation du programme « Youth Opportunities Program », visant à améliorer les perspectives d'emploi des jeunes. Les bénéficiaires du NUSAF sont-ils différents de la population en général ? Si tel est le cas, comment établir un contrefactuel valable ?

Le Gouvernement et l'équipe de recherche s'attendaient à ce qu'il y ait d'importantes différences entre les bénéficiaires du NUSAF et la population en général. Les candidats étaient en effet supposés former des groupes et présenter des propositions. Cela signifiait qu'ils devaient être quelque peu instruits, et donc plus aisés. En outre, ceux qui soumettaient des propositions devaient être résolus à s'engager dans les affaires et étaient donc probablement très motivés. Il s'agit là d'une caractéristique qui n'est pas facilement mesurable.

Pour vérifier les différences potentielles, le NUSAF a observé les caractéristiques des participants au programme collectées au départ et les a comparées à celles d'autres jeunes interrogés à peu près à la même période. Il a trouvé que les membres du programme avaient beaucoup plus de biens et avaient un bien meilleur niveau d'éducation que la population en général. De plus, les femmes étaient largement sous-représentées dans le programme (33 %) comparées à la population en général (51 %). Les ménages considérés dans l'étude avaient cinq fois plus de chances d'avoir une radio ou une bicyclette et trois fois plus de chances d'avoir un téléphone portable ou du bétail que la population en général. Il y avait également des disparités entre le niveau d'éducation des participants au programme et celui de la population en général.

En outre, la comparaison des taux de pauvreté dans la population en général avec ceux parmi les participants au programme a révélé des différences frappantes : au moins 50 % des participants au programme étaient au-dessus du seuil de pauvreté. Ainsi, que l'on considère les différences relatives ou absolues entre la population en général et les participants au programme, il apparaissait clairement qu'en moyenne les candidats au programme faisaient partie d'une couche socioéconomique plus élevée par rapport à un échantillon représentatif des jeunes de la région.

Les différences entre les groupes ont mis en évidence les raisons pour lesquelles il était nécessaire d'être prudent dans la conduite d'une évaluation d'impact. En utilisant la population en général comme contrefactuel, on surestimerait grandement l'effet du programme, étant donné qu'il y avait déjà des différences majeures entre les échantillons de population, même en absence de programme.

Pour identifier un contrefactuel valable, l'équipe d'évaluation pourrait profiter du fait que les demandes de participation au programme étaient très nombreuses mais que les ressources financières disponibles étaient maigres. Étant donné les problèmes liés à la constitution d'un groupe témoin valable, décrits plus haut, et les fonds insuffisants pour que toute personne remplissant les conditions requises puisse participer au programme, il a été décidé d'utiliser les méthodes randomisées pour sélectionner les sujets du groupe témoin.

Source : Blattman, Fiala et Martinez (2011).

À lire

Gertler, P., Martinez, S., Premand, P., Rawlings, L., et Vermeersch, C., 2011,
L'évaluation d'impact en pratique, Washington, D.C., La Banque mondiale, voir chapitre 3,
http://siteresources.worldbank.org/INTHDOFFICE/Resources/IEP_Ebook_FRENCH.pdf

Notes

NOTE 6 : Choisir une méthode d'évaluation d'impact appropriée

La présente note a pour but de présenter aux praticiens un aperçu des différents outils disponibles pour procéder à une évaluation d'impact et de donner des orientations sur les outils qui peuvent être les plus adaptés à un projet particulier. Nous présenterons six méthodes couramment utilisées dans les évaluations d'impact, classées en fonction de leur utilité pour constituer un contrefactuel avec un biais minimum dans l'estimation. Le choix d'une méthode d'évaluation d'impact dépendra non seulement de la qualité théorique de la méthode mais aussi du contexte dans lequel le programme est mis en œuvre. Les responsables de programmes doivent donc participer à la conception de la procédure d'évaluation pour s'assurer que l'évaluation répond aux besoins et qu'elle est adaptée au contexte.

Opérer un choix parmi les méthodes d'évaluation d'impact

Chaque technique d'évaluation d'impact diffère en termes de circonstances dans lesquelles elle est le mieux appliquée ; les évaluations ne conviennent pas toutes à tous les contextes. Les caractéristiques de notre programme et les circonstances dans lesquelles il se déroule guideront donc notre choix parmi les méthodes d'évaluation d'impact. Comme Gertler *et al.* (2011, p. 143 à 149) l'illustrent, nous devons notamment prendre en considération le moment, la couverture, la cible et les ressources :

Le moment

La mise en œuvre du programme a-t-elle déjà commencé ? La question essentielle à se poser ici est si l'évaluation d'impact peut être intégrée à la conception du programme. Comme nous l'expliquerons plus en détail après, lorsqu'une évaluation d'impact est prévue dès le début du programme, la qualité de l'évaluation est bien meilleure, et il est possible de choisir une méthode parmi un éventail beaucoup plus large.

La couverture

Les personnes remplissant les conditions requises peuvent-elles toutes bénéficier du programme ? Idéalement, nous aimerions que tous les jeunes dans le besoin en profitent. C'est plus facile avec certains types de programmes qu'avec d'autres. Si le programme ne requiert pas beaucoup de ressources (comme l'ouverture de comptes d'épargne pour les mineurs) ou qu'il est fourni à travers les organes d'information (campagne médiatique télévisée ou radiophonique destinée à améliorer les compétences financières), alors il se peut que nous ne voulions pas – ou même que nous ne puissions pas – exclure qui que ce soit du programme. Dans la plupart des cas, cependant, nous ne disposons pas de ressources suffisantes pour offrir des programmes visant à améliorer les moyens de subsistance des jeunes à tous ceux qui remplissent les conditions requises, ce qui nous oblige à choisir qui, parmi les jeunes admissibles, bénéficiera du programme et qui n'en bénéficiera pas. S'il peut être frustrant, au niveau de l'élaboration des programmes, de ne pas pouvoir aider tous les jeunes, le nombre très élevé de demandes offre la possibilité de constituer un groupe témoin et de conduire des évaluations de qualité sur l'impact du programme.

Le ciblage

Comment les bénéficiaires du programme sont-ils sélectionnés ? À moins d'être en mesure de proposer le programme à tous les jeunes qui remplissent les conditions requises, nous devons sélectionner des personnes ou des groupes à l'aide des moyens suivants :

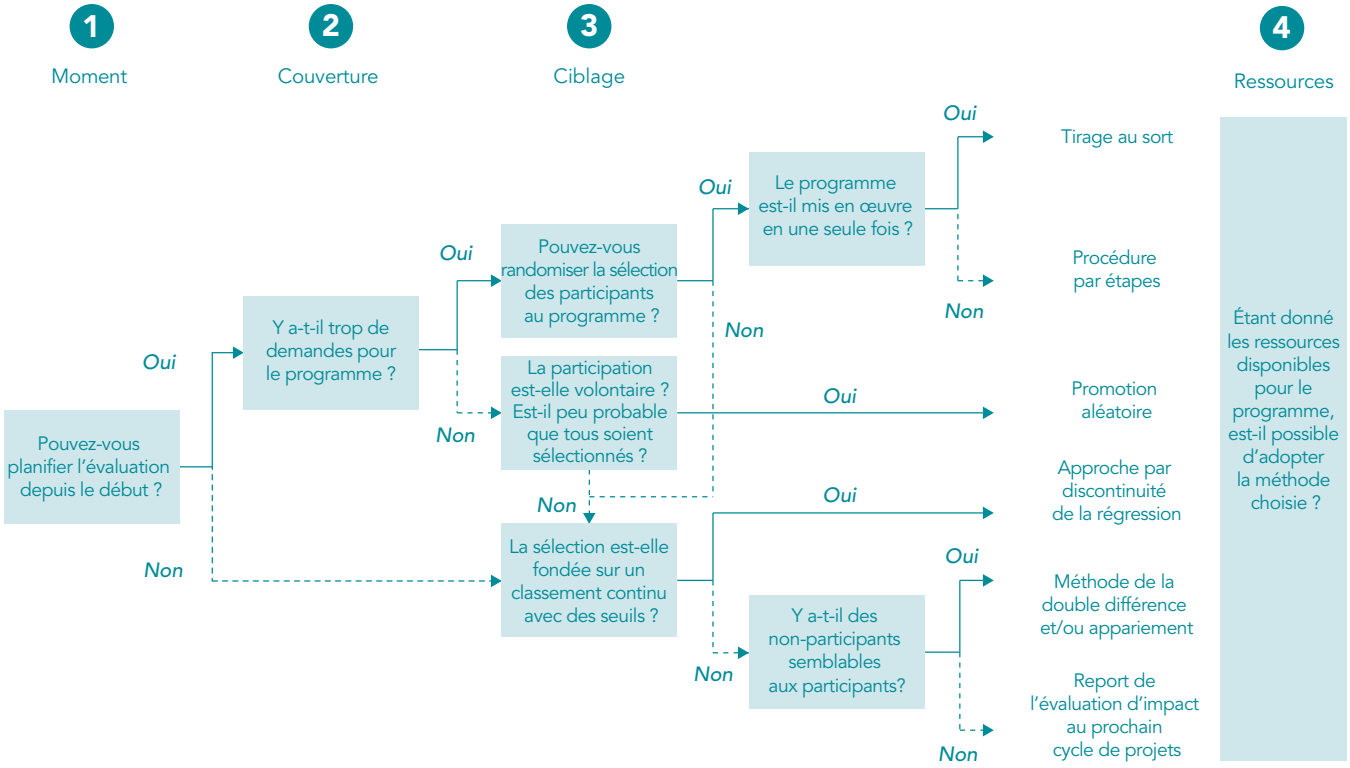
1. **La répartition aléatoire**, qui consiste à donner à chaque personne ou groupe les mêmes chances de bénéficier du programme, par exemple en tirant au sort des noms pour déterminer qui suivra la formation professionnelle et qui sera inscrit sur la liste d'attente.
2. **Le classement selon des critères d'admissibilité**, qui détermine l'admissibilité selon des critères clairement établis qui permettent de fixer une limite ou un seuil. Il s'agit par exemple d'attribuer des bourses d'études en fonction des résultats obtenus à des tests de contrôle ou d'offrir une formation en fonction des niveaux de revenu.
3. **Le ciblage sélectif.** Il arrive qu'il n'y ait pas de critère clairement défini expliquant pourquoi une personne ou un groupe est sélectionné plutôt qu'un autre, ce qui, au lieu d'assurer l'équité dans la sélection, entraîne un biais dans la sélection des participants. Les sélections dans l'ordre des demandes, en fonction de facteurs politiques ou pour des raisons pratiques sont des exemples de méthodes de sélection subjectives par nature.

Ressources

Le programme est-il doté des ressources suffisantes pour qu'une évaluation d'impact soit réalisée ? La technique d'évaluation d'impact sera choisie en fonction de différents critères, en termes de taille de l'échantillon, de collecte des données, de complexité de l'analyse statistique et de coûts. Même lorsque nous optons pour une méthode qui conviendrait au contexte dans lequel nous opérons, il se peut que nous ne puissions pas l'adopter faute de ressources.

Nous devrions garder à l'esprit les quatre questions susmentionnées lorsque nous examinons différentes techniques d'évaluation d'impact. La réponse apportée à ces questions nous permettra de déterminer laquelle des six méthodes est la plus adaptée à notre contexte (voir figure 6.1.). Nous nous pencherons ensuite sur les méthodes d'évaluation elles-mêmes.

FIGURE 6.1 Arbre de la décision permettant de choisir une technique d'évaluation d'impact



Sources : arbre conçu à partir de GAO (1991, p. 69) ; Duflo, Glennerster et Kremer (2006, p. 24 à 27) ; et Gertler *et al.* (2011, p. 148).

[Définition]

Un **essai contrôlé randomisé** est une étude dans laquelle les personnes sont assignées de manière aléatoire (au hasard uniquement) à un groupe qui bénéficiera d'un traitement, telle que la participation à une intervention spécifique.

Un **échantillon** est un sous-ensemble d'une population. S'il est habituellement impossible ou difficilement réalisable de recueillir des informations sur l'ensemble de la population d'intérêt, il est en revanche possible de recueillir des informations sur un sous-ensemble de taille plus facilement gérable. Si le sous-ensemble est bien choisi, il est possible d'extrapoler à partir des résultats obtenus ou d'en tirer des conclusions qui vaudront pour l'ensemble de la population.

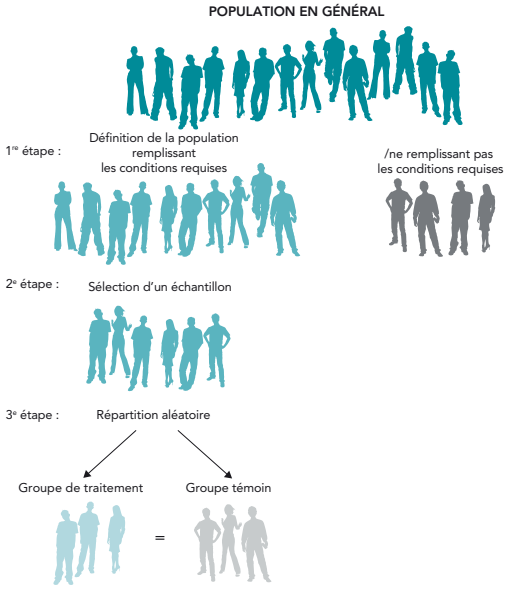
1re méthode : la procédure par tirage au sort

La procédure par tirage au sort est un moyen simple et transparent de répartir des personnes entre un groupe qui bénéficiera de nos services (groupe de traitement) et un autre qui n'en bénéficiera pas (groupe témoin). C'est la méthode utilisée pour concevoir les essais contrôlés randomisés. Si un échantillon suffisamment important de personnes issues de la même population d'intérêt est réparti de manière aléatoire entre les deux groupes, alors les deux groupes possèderont, en moyenne, des caractéristiques observables (âge, sexe, taille, niveau de scolarité, etc.) et non observables (motivations, état d'esprit) semblables ; il s'agit là d'une régularité statistique. Avec l'aléation, la différence observée entre les résultats des deux groupes au terme du programme peut être attribuée à l'intervention, car tous les autres facteurs qui pourraient influencer les résultats sont, en moyenne, équivalents. Les tirages au sort sont considérés comme le type d'évaluation d'impact le plus rigoureux, et en général c'est, pour les donateurs, les acteurs concernés et les gouvernements, la méthode qui donne les résultats les plus fiables.

Comment procède-t-on ?

Un tirage au sort comporte trois étapes (voir figure 6.2.).

FIGURE 6.2 Étapes d'un tirage au sort



1re étape : définition de la population remplissant les conditions requises

La première étape dans un essai contrôlé randomisé consiste à constituer un groupe de personnes remplissant les conditions requises pour participer à un programme. Si un expert médical étudie l'effet d'un médicament sur une maladie infantile, il essaiera de constituer un groupe d'enfants et n'inscrira pas d'adultes ou de personnes âgées au programme. De la même façon, un programme destiné à renforcer les moyens de subsistance des jeunes peut cibler les jeunes citadins d'une certaine tranche d'âge qui se trouvent dans la rue, et ne visera donc pas les adultes ou les jeunes vivant en zone rurale. Ce qui est important à ce stade, c'est de définir clairement des critères en toute transparence (âge, sexe, niveau de revenu, situation au regard de l'emploi, etc.) et de pouvoir communiquer qui sera au bénéfice du programme et qui ne le sera pas.

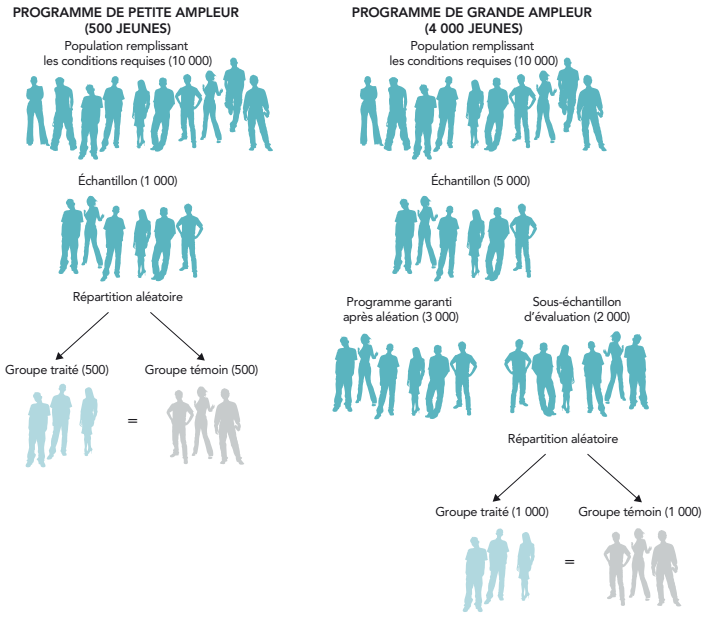
2e étape : sélection d'un échantillon d'évaluation

Pour évaluer une intervention, nous n'avons pas besoin de faire passer des tests à tous ceux qui participeront à l'intervention. Il nous faut seulement choisir un groupe représentatif de personnes qui soit suffisamment nombreux pour que l'évaluation puisse être conduite ; c'est ce qu'on appelle l'*échantillon* (voir note 7 pour plus de détails sur la manière de déterminer l'échantillon et sa taille). Ce seront les personnes sur lesquelles nous collecterons des données.

Le choix de l'échantillon d'évaluation peut se faire de deux façons selon que le programme est de petite ou de grande ampleur. Dans le cas d'un programme de petite ampleur, on peut trouver 10 000 personnes remplissant les conditions requises, par exemple être un jeune citadin âgé de 16 à 24 ans et vivre dans la rue. Avec le budget dont le programme est doté, on peut aider 500 de ces jeunes. Dans l'idéal, le groupe témoin sera aussi nombreux que le groupe de traitement. On devra donc sélectionner pour le programme et l'évaluation 1 000 jeunes des rues sur les 10 000 potentiels (voir figure 6.3., illustration de gauche).

Les programmes de grande ampleur peuvent parfois viser un groupe de personnes plus nombreux que la taille de l'échantillon nécessaire à l'évaluation. Si le programme peut bénéficier à 4 000 jeunes, il n'est pas nécessaire de trouver 4 000 autres jeunes pour procéder à la comparaison. Seuls 1 000 peuvent suffire. On peut alors constituer un échantillon de 5 000 jeunes parmi les 10 000 recensés dans la population totale. Parmi ces 5 000 jeunes, 3 000 peuvent être assurés de participer au programme. Les 2 000 restants seront répartis aléatoirement entre deux groupes, le groupe de traitement et le groupe témoin (voir figure 6.3., illustration de droite).

FIGURE 6.3 Choisir des échantillons pour des programmes de petite ou de grande ampleur



Pour que la sélection soit représentative de la population totale des 10 000 jeunes des rues remplissant les conditions requises, l'échantillon (soit 1 000 jeunes dans le premier cas et 5 000 dans le second) devrait être sélectionné aléatoirement parmi la population admissible. Avec une sélection aléatoire, les participants au programme possèderont, en moyenne, des caractéristiques semblables à celles de l'ensemble de la population admissible. Même si l'étude ne porte que sur un nombre limité de jeunes, l'impact potentiel du programme peut être généralisé à l'ensemble de la population admissible, soit dans ce cas 10 000 jeunes.

[Conseil]

Un des moyens d'obtenir un échantillon randomisé de jeunes consiste à se procurer une liste de tous les jeunes des rues dressée à partir d'un recensement, de la liste électorale ou de toute autre base de données, et de procéder à une sélection aléatoire à partir de cette liste. Si ce n'est pas possible, il faut cibler aléatoirement les zones où les jeunes des rues se réunissent, comme un centre urbain, pour produire un échantillon randomisé. Si l'on sait que des jeunes se retrouvent dans une cinquantaine de centres urbains autour de la ville ou à travers le pays, le fait de sélectionner aléatoirement ces centres, puis de sélectionner une partie des jeunes fréquentant ces centres pour qu'ils participent à l'étude permettra probablement d'assurer une sélection de jeunes avec un biais minime. Nous reparlerons plus en détail de l'échantillonnage dans la note 7.

3e étape : répartition aléatoire

L'étape suivante consiste à répartir l'échantillon de jeunes sélectionnés entre le groupe de traitement et le groupe témoin, plus ou moins équivalents en termes de taille. Dans les essais contrôlés randomisés, tous les jeunes ont les mêmes chances de suivre le programme. On peut procéder à l'aléation par des techniques traditionnelles comme en lançant une pièce de monnaie en l'air (pile ou face), en jetant des dés ou en tirant au sort des noms. L'aléation peut avoir lieu publiquement, si l'on préfère et si l'échantillon est relativement peu nombreux (tirer au sort 2 000 noms, par exemple, ne serait pas très pratique). Il est également possible – et préférable si le nombre de personnes est élevé – de procéder à l'aléation à l'aide d'un logiciel informatique, tel que MS Excel. L'aléation peut être utilisée à plusieurs niveaux (voir encadré 6.1.). En répartissant aléatoirement notre échantillon entre le groupe de traitement et le groupe témoin, nous sélectionnons des participants de manière équitable et nous mettons aussi en place un contrefactuel valable : si la taille de l'échantillon est suffisamment grande, les sujets du groupe de traitement ont, en moyenne, les mêmes caractéristiques observables et non observables que ceux du groupe témoin.

ENCADRÉ 6.1 Niveaux d'aléation

L'aléation peut se faire au niveau des personnes, des groupes ou des communautés, selon les besoins du programme.

Au niveau des personnes. L'aléation au niveau des personnes est préférable dans le cas de projets où les résultats seront mesurés pour chaque participant. Cette méthode peut poser des problèmes, tels que des effets de propagation, qui surviennent lorsque les sujets du groupe témoin bénéficient d'une partie du traitement par des moyens informels. Par exemple, les jeunes qui ont reçu une formation ou autres informations dans le cadre de notre programme partagent leurs connaissances ou leurs ressources avec leurs amis membres du groupe témoin.

Au niveau des groupes. L'aléation au niveau des personnes n'est pas toujours possible ou souhaitable. Si l'on ne dispose pas tout de suite d'une liste de noms, ou qu'il y a des chances que les personnes sélectionnées pour constituer le groupe témoin bénéficient de toute façon du programme, alors l'aléation au niveau des groupes est peut-être préférable. Cette méthode convient tout particulièrement dans le cas de programmes qui visent des groupes, tels que les écoles, les centres de formation professionnelle, les maisons des jeunes, etc. Dans ce cas, les groupes de personnes sont répartis aléatoirement entre cohortes de traitement et cohortes témoins. Tous les sujets du groupe de traitement bénéficieront de la même intervention. L'aléation au niveau des groupes peut contribuer à réduire les effets de propagation et peut être plus facile à utiliser que l'aléation au niveau des personnes. On peut aussi procéder à une aléation au niveau des sous-groupes, tels que les classes d'une école.

Au niveau des villages/des communautés. Dans le cadre de certains programmes, on peut également procéder à une aléation au niveau des villages, des quartiers, des communautés ou même des districts, lorsque des activités sont réalisées à ce niveau ou qu'il y a des risques d'effets de propagation au-delà du groupe. Par exemple, si le district d'intérêt comprend 100 villages et que nous ne disposons pas des ressources suffisantes pour travailler avec tous ces villages, nous pouvons sélectionner de manière aléatoire 50 villages pour travailler avec eux et garder les 50 autres pour former le groupe témoin. Tous les jeunes vivant dans les villages de traitement rempliraient alors les conditions requises pour participer au programme.

(suite page suivante)

ENCADRÉ 6.1 (SUITE) Niveaux d'aléation

Une aléation à un niveau plus élevé, dans le cas d'une intervention à ce niveau, peut, bien qu'elle puisse contribuer à réduire les effets de propagation non désirés, être également problématique pour les raison suivantes :

- Plus le niveau d'aléation est élevé, plus le nombre d'observations qui peuvent être comparées entre elles sera petit. La réalisation d'entretiens avec un certain nombre de personnes par zone peut permettre d'atténuer ce problème.
- La taille de l'échantillon d'évaluation est proportionnelle à l'ampleur de l'intervention, ce qui a des répercussions sur le coût de l'évaluation.
- Les unités à des niveaux plus élevés risquent davantage de subir les influences extérieures au fil du temps, ce qui peut rendre plus difficile la comparaison du groupe de traitement avec le groupe témoin et compromettre la validité interne de l'évaluation.

Les responsables de programmes devraient par conséquent déterminer à quelle échelle minimum il est possible de mettre en œuvre un programme et à quel niveau il convient de sélectionner aléatoirement les participants.

Quand peut-on utiliser une procédure par tirage au sort ?

On utilise l'évaluation par tirage au sort lorsque l'évaluation est prévue avant la mise en œuvre du programme (évaluation prospective) et que le programme peut bénéficier seulement à une partie des participants remplissant les conditions requises. Tant que les contraintes en termes de ressources empêchent d'étendre le programme à l'ensemble de la population admissible, il n'y a pas de problèmes éthiques à constituer un groupe témoin, car une partie de la population sera nécessairement exclue du programme. Dès lors, on peut garder les groupes témoins pour mesurer les effets à court, moyen et long terme du programme (Gertler *et al.* 2011).

Dans toute évaluation prospective, il faudra collecter de nouvelles données, ce qui comporte un coût. Il faudra au minimum réaliser une étude finale (dont on reparlera en détail dans la note 7) sur les jeunes formant le groupe de traitement et le groupe témoin. Dans bien des cas, il sera également nécessaire de mener une enquête initiale. Malgré les coûts liés à la collecte de nouvelles données, l'option la moins chère pour une évaluation peut être un simple tirage au sort, car il nécessite moins d'enquêtes et un nombre plus restreint de personnes interrogées.

[Définition]

Une **évaluation prospective** a pour objet de suivre les participants après la fin du programme. Ces études doivent donc être planifiées lors de l'élaboration du programme.

On entend par **évaluations rétrospectives** les évaluations qui visent à examiner les participants à des programmes qui ont déjà été mis en œuvre ou qui sont même terminés.

Avantages

- Le tirage au sort est la méthode la plus fiable pour constituer un contrefactuel car il permet de former un groupe témoin correspondant très bien au groupe de traitement (fondé sur moins d'hypothèses que d'autres méthodes). Il est donc considéré comme la meilleure méthode pour évaluer l'impact.
- C'est de loin la méthode d'évaluation la plus simple sur le plan analytique. Dans un essai randomisé, l'impact du programme est simplement la différence moyenne entre les résultats obtenus dans le groupe de traitement et ceux obtenus dans le groupe témoin.
- Il permet aux communautés d'être directement impliquées dans le processus de sélection et d'assurer une allocation équitable et transparente des bénéfices.
- Étant donné qu'il est prévu dès le lancement du programme, il peut être conçu non seulement pour mesurer l'impact moyen du programme, mais aussi pour comparer l'efficacité des différents éléments et des différentes durées du programme, etc.
- Il est facile de l'organiser et de le transmettre aux exécuteurs du programme.

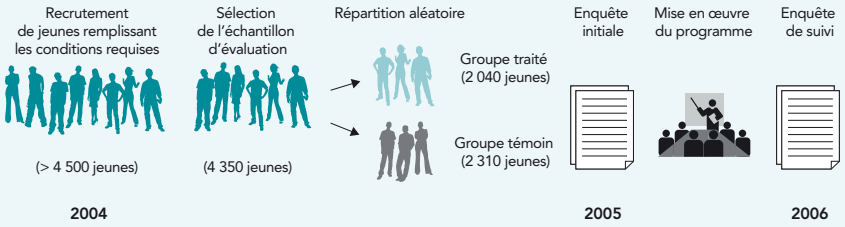
Inconvénients

- Il faut constituer un groupe témoin qui sera exclu du programme pour la durée de l'évaluation d'impact.
- Les organisations doivent faire en sorte que les partenaires et les acteurs locaux consentent à adopter cette méthode.
- La validité interne d'un tirage au sort dépend de l'efficacité de l'aléation et de son utilisation durant toute la durée de l'étude, ce qui n'est pas assuré. Cette condition n'est pas forcément garantie si l'aléation est faite de manière incorrecte, que le groupe de traitement ou le groupe témoin ne respecte pas son statut (c'est-à-dire, si les personnes traitées ne suivent pas le programme ou que les personnes témoins suivent le programme), que les participants abandonnent l'étude avant la fin ou qu'il y a des effets de propagation.

L'encadré 6.2. nous donne un exemple de procédure par tirage au sort.

ENCADRÉ 6.2 Exemple de procédure par tirage au sort

Attanasio, Kugler et Meghir (2009) ont utilisé une procédure par tirage au sort pour étudier le programme d'emploi des jeunes « Jóvenes en Acción » mis en œuvre en Colombie pour fournir aux jeunes de 18 à 25 ans issus de la couche socioéconomique la plus basse de la population une formation théorique de trois mois suivie d'une formation en cours d'emploi. Les formateurs ont reçu l'instruction de recruter plus de candidats qu'ils n'avaient de place dans leurs cours, au cas où certaines personnes se désisteraient. Les participants ont ensuite été sélectionnés de manière aléatoire à partir du groupe de candidats recrutés, et les jeunes qui restaient ont été inscrits sur une liste d'attente et utilisés comme groupe témoin.



Attanasio et ses collègues craignaient que, malgré l'aléation, le groupe de traitement puisse être différent du groupe témoin sur des aspects que les chercheurs ne pourraient pas contrôler. En reprenant les données initiales, ils ont vérifié la comparabilité des deux groupes et trouvé que, en moyenne, le groupe de traitement avait fréquenté l'école trois mois de plus que le groupe témoin et qu'il comptait environ 5 % de jeunes femmes de plus que le groupe témoin. Ils ont estimé qu'aucune de ces caractéristiques n'influencerait de manière significative les résultats du traitement.

Dans l'ensemble, les résultats ont été prometteurs. En moyenne, ceux qui avaient suivi le programme avaient plus de chances d'occuper un emploi rémunéré dans le secteur formel, d'avoir des revenus plus élevés et de garder leur emploi plus longtemps que les sujets du groupe témoin. Les effets ont été généralement plus marqués chez les femmes que chez les hommes.

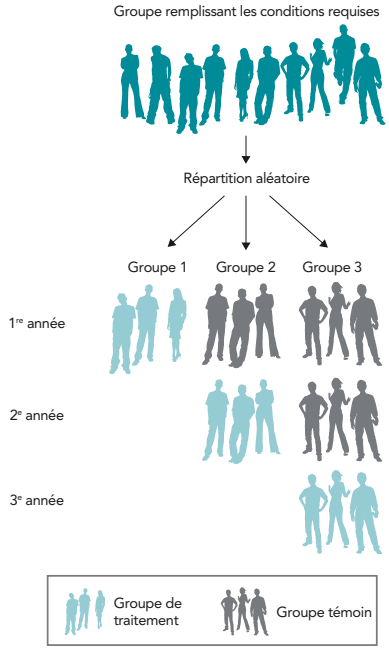
2e méthode : procédure randomisée par étapes

Il est parfois impossible de constituer un groupe témoin dont les membres ne bénéficieront jamais du programme. Lorsque de nombreux programmes sont réalisés dans une communauté pendant des années, il peut être difficile, tant sur le plan politique qu'au niveau de la programmation, de ne jamais en faire profiter un groupe de personnes dans le besoin. La procédure par étapes est une variante du tirage au sort. Elle s'applique aux programmes qui sont mis en œuvre par étapes et elle utilise le flux naturel de personnes pour constituer le groupe de traitement et le groupe témoin.

Comment procède-t-on ?

La principale différence entre une procédure par étapes et une procédure par tirage au sort est la méthode de répartition des personnes entre le groupe de traitement et le groupe témoin. Quand une intervention est menée en plusieurs phases, la procédure par étapes offre à chaque personne ou groupe remplissant les conditions requises les mêmes chances de bénéficier du programme à chaque phase de sa mise en œuvre. Un groupe de jeunes est sélectionné aléatoirement pour recevoir le traitement au cours de la première phase, alors qu'un autre groupe est sélectionné pour suivre le programme au cours de la deuxième phase, un troisième groupe au cours de la troisième phase, etc. Les groupes qui figurent sur la liste d'attente peuvent servir de groupe témoin jusqu'à ce qu'ils suivent le programme (voir figure 6.4.).

FIGURE 6.4 Groupe de traitement et groupe témoin dans la procédure par étapes



Note : Le traitement ne doit pas nécessairement être arrêté pour la conduite de l'évaluation. Certaines interventions, une fois en cours, continueront d'être mises en œuvre. Cependant, de nombreux programmes, tels que les formations, sont offerts sur une période de temps limitée.

Par exemple, une ONG peut disposer du budget pour former 1 500 jeunes mais ne pas avoir les capacités pour assurer l'ensemble de la formation tout de suite. Elle peut alors choisir de former 500 personnes par an pendant trois ans. Si elle peut trouver l'ensemble des 1 500 participants dès le début, la procédure randomisée par étapes sera peut-être la meilleure méthode d'évaluation.

Les 1 500 jeunes sont répartis aléatoirement entre trois groupes. La première année, tandis que le groupe 1 suit la formation, les groupes 2 et 3 sont inscrits sur une liste d'attente et peuvent servir de groupe témoin. La deuxième année, seul le groupe 3 reste dans le groupe témoin. À la fin de la troisième année, les trois groupes auront tous suivi la formation.

Comme les sujets sont sélectionnés au hasard pour constituer les différents groupes, il est possible de comparer ceux qui bénéficient du traitement en premier avec ceux qui en bénéficieront en dernier. Cependant, étant donné que tous suivront finalement le programme et que la dernière année il n'y aura plus de groupe témoin, la procédure par étapes ne convient généralement pas pour évaluer l'impact à long terme d'un programme. Même dans le cas de programmes à long terme et de grande ampleur, il sera difficile de demander aux participants d'attendre pendant trois ou quatre ans ; on ne pourra donc observer les résultats que sur une durée souvent limitée à un ou deux ans.

Quand peut-on utiliser une procédure par étapes ?

Comme lors d'une procédure par tirage au sort, une évaluation par étapes est prospective et n'est envisageable que s'il y a une forte demande et que l'on a la possibilité de répartir aléatoirement les participants entre le groupe de traitement et le groupe témoin. La procédure par étapes est mieux adaptée que la procédure par tirage au sort aux programmes de grande envergure, où les interventions doivent se dérouler sur un certain nombre d'années. De même qu'il est nécessaire de prévoir, dans le cadre de la procédure par étapes, un plan de déroulement, il est aussi nécessaire de constituer une équipe qui sera en mesure de suivre le déroulement du programme durant toute sa durée.

Les procédures par étapes ne diffèrent pas sensiblement des procédures par tirage au sort en termes de données ou de coûts. Dans de nombreux cas il faudra mener une enquête finale ainsi qu'une enquête initiale. Différence majeure, les coûts de mise en œuvre du programme risquent d'augmenter, car des ressources seront nécessaires pour assurer la conformité du déroulement du programme avec l'évaluation.

[Conseil]

Avec une procédure par étapes, il est essentiel d'attendre suffisamment de temps entre chacune des phases afin de pouvoir observer les effets du programme. Si un responsable de programme estime qu'il faut attendre deux ans avant que le programme donne des résultats, la période de temps entre la première et la dernière phase doit être d'au moins deux ans. Cette procédure ne convient pas forcément pour les programmes de petite ampleur ou mis en œuvre à court terme.

Avantages

- Les procédures par étapes produisent un contrefactuel fiable, assurent une sélection équitable et transparente, et permettent de comparer les effets de programmes déclinés différemment.
- La méthode convient au déroulement naturel de nombreux programmes.
- Comme tout le monde suit finalement le programme, les évaluations par étapes sont avantageuses sur le plan politique.

Inconvénients

- Comme avec la procédure par tirage au sort, il est difficile de garantir une bonne aléation et de conserver le groupe de traitement et le groupe témoin dans la durée.
- Il se peut que les participants ne veuillent pas attendre de suivre le programme. Si tel est le cas, le risque est qu'ils changent d'attitude entre-temps et qu'ils ne constituent plus un bon groupe témoin. Par exemple, il se peut qu'ils arrêtent de rechercher un emploi en attendant de suivre le programme.
- Avec la procédure par étapes, on ne peut pas estimer l'impact du programme à long terme.
- Pour appliquer cette méthode, il faut définir clairement une stratégie de déroulement, ce qui peut avoir des implications sur le plan de la mise en œuvre.

Voir encadré 6.3. pour un exemple de procédure randomisée par étapes.

ENCADRÉ 6.3 Exemple de procédure randomisée par étapes

Le projet d'autonomisation économique des adolescentes, lancé au Libéria par la Banque mondiale, propose une formation de six mois et des activités de suivi durant six mois autour de deux programmes différents : 1) formation professionnelle en vue d'un emploi salarié et accompagnement vers l'emploi ; 2) renforcement des compétences en création d'entreprise et établissement de contacts avec des acteurs de la micro-finance. Un encadrement est aussi prévu pour tous les bénéficiaires à partir du troisième mois de formation.

Pour mesurer l'impact de ces programmes, la Banque mondiale a opté pour une évaluation par étapes, qui permet d'avoir une évaluation randomisée de qualité et de proposer finalement le programme à toutes les filles auxquelles on a promis une formation. L'évaluation a tiré parti du déroulement naturel du programme et des contraintes liées à la mise en œuvre, qui ne permettaient pas de former toutes les filles en même temps.

À l'issue de l'enquête initiale, 1 273 participantes ont été assignées de manière aléatoire au groupe de traitement (et ont suivi une formation au cours du premier cycle du programme, en 2010), et 843 au groupe témoin (et ont suivi une formation lors du deuxième cycle du programme, en 2011). L'enquête de suivi a été conduite à la fin de chaque cycle, complétée par des sondages sur la qualité du programme, qui sont destinés à recueillir l'avis des participants à l'issue du parcours, sur la formation, le contenu, la pédagogie et les formateurs.

Comme le programme et l'évaluation ciblaient les filles qui avaient manifesté leur intérêt pour la formation, les résultats de l'évaluation ne peuvent pas être généralisés à l'ensemble des jeunes femmes de la population. L'évaluation nous aide à comprendre l'impact de la formation sur les personnes qui ont choisi de suivre une formation et d'être encadrées dans un emploi salarié ou lors de la création de leur entreprise.

Sources : Banque mondiale (2008) ; Muzi (2011).

3e méthode : promotion aléatoire

Il peut y avoir des cas où il n'est pas possible ou souhaitable d'exclure de potentiels bénéficiaires, soit parce que la participation est volontaire et que toute personne peut s'inscrire si elle le souhaite, soit parce que le programme est doté d'un budget suffisant pour que l'ensemble de la population de jeunes remplissant les conditions requises puisse bénéficier immédiatement de l'intervention. Dans ces cas, la méthode de la promotion aléatoire (également appelée modèle d'encouragement) peut convenir.

Comment procède-t-on ?

Avec la promotion aléatoire, on identifie la population remplissant les conditions requises et on choisit un échantillon tout comme dans les procédures par tirage au sort ou par étapes. Mais la méthode d'aléation diffère. Lorsqu'il n'est pas possible de répartir aléatoirement des jeunes entre un groupe qui bénéficie d'un programme et un groupe qui n'en bénéficie pas, il peut être possible en revanche de promouvoir de manière aléatoire le programme. C'est-à-dire, plutôt que de sélectionner aléatoirement les personnes qui *profiteront* des avantages et des services offerts, nous sélectionnons aléatoirement les personnes qui sont *encouragées à profiter* de ces avantages.

La promotion aléatoire est fondée sur l'hypothèse que pour beaucoup de programmes il y aura trois groupes de bénéficiaires potentiels :

- Les jeunes qui ne s'inscrivent jamais ;
- Les jeunes qui s'inscrivent toujours ;
- Les jeunes qui ne s'inscrivent que s'ils y sont encouragés.

Quoique le programme offre, comptes d'épargne gratuits, formation professionnelle ou campagnes médiatiques destinées à améliorer les compétences financières, il est généralement peu probable que tous les jeunes remplissant les conditions requises veuillent y participer. Certains

peuvent tout simplement douter de l'intervention, d'autres peuvent manquer de temps ou avoir des problèmes de transport, d'autres encore peuvent ne pas avoir eu connaissance du programme.

L'encouragement aléatoire peut prendre différentes formes. Dans le cas de comptes d'épargne gratuits, nous pouvons faire la publicité de l'initiative dans des écoles sélectionnées au hasard. Pour un programme de formation, nous pourrions recruter un travailleur social pour qu'il visite au hasard certains ménages comprenant des jeunes au chômage, qu'il décrive le programme et qu'il propose d'inscrire les jeunes sur place. Dans le cas d'une campagne médiatique destinée à améliorer les compétences financières, nous pourrions envoyer des SMS (textos) à une partie du public cible sélectionnée au hasard et pas à l'autre. Dans tous les cas, il y aura toujours des personnes dans le groupe encouragé qui ne suivront pas le programme, tout comme il y aura des personnes dans le groupe non encouragé qui le suivront. Mais l'idée est que si l'encouragement est efficace, alors le taux d'inscription dans le groupe encouragé devrait être plus élevé que celui dans le groupe qui n'a pas bénéficié de la promotion. Et si la promotion a été menée de manière aléatoire, alors le groupe encouragé et le groupe non encouragé possèdent, en moyenne, les mêmes caractéristiques, ce qui permet de déterminer les effets de causalité.

L'évaluation au moyen de la promotion aléatoire peut convenir pour :

- *les programmes qui distribuent des offres de formation ;*
- *les programmes qui encouragent les jeunes à ouvrir des comptes d'épargne ;*
- *les interventions qui influencent les campagnes médiatiques.*

Malheureusement, nous ne pouvons pas simplement comparer les résultats des personnes qui ont participé au programme avec les résultats de celles qui n'y ont pas participé. Comme il a été vu dans la note 5, les personnes qui décident de suivre un programme sont presque toujours différentes de celles qui ne le suivent pas, et beaucoup de ces différences ne sont peut-être pas observables ou mesurables. Même si la promotion est aléatoire, la participation au programme ne sera pas due au hasard, alors comparer des participants avec des non-participants reviendrait à comparer des pommes et des oranges.

Ce que nous pouvons faire en revanche, c'est comparer les résultats de tous les jeunes qui ont été encouragés avec les résultats de ceux qui ne l'ont pas été (voir figure 6.5.). Considérons par exemple un programme de formation professionnelle auquel ont participé 30 % des jeunes remplissant les conditions requises, issus du groupe non encouragé, et 80 % des jeunes admissibles issus du groupe encouragé (Gertler et al. 2011). Un an après la fin du programme, nous observons un revenu mensuel de 60 dollars en moyenne pour le groupe non encouragé et de 100 dollars pour le groupe encouragé.

FIGURE 6.5 Estimation de l'impact dans le cadre d'une promotion aléatoire

	Groupe non encouragé	Groupe encouragé	Changements observés
Inscription (% de la population remplissant les conditions requises	30%	80%	50%
1er type : Ne s'inscrit jamais			
2e type : S'inscrit toujours			
3e type : Ne s'inscrit que si encouragé			
Résultat moyen (revenu mensuel)	USD 60	USD 100	USD 40
Effet de causalité			**USD 80** (= USD 40/0.5)

Personnes qui s'inscrivent dans les deux scénarios - Source: d'après Gertler et al. (2011, p. 75).

Étant donné que la promotion se fait de manière aléatoire, le groupe encouragé et le groupe non encouragé présentent, en moyenne, des caractéristiques semblables. Ainsi, la différence que nous observons entre les résultats moyens obtenus par les deux groupes (à savoir 40 dollars) peut être attribuée au fait que dans le groupe de personnes qui s'inscrivent, seules les personnes qui y ont été encouragées suivent le programme. Bien que nous ne puissions pas distinguer directement ces personnes de celles qui s'inscrivent toujours, nous savons que leur part représente, sur l'ensemble de la population, la différence qui apparaît dans les taux d'inscription (50 % ou 0,5). Ainsi, l'impact moyen du programme sur les personnes qui y ont participé parce qu'elles y avaient été encouragées est de 40/0.5, soit 80 dollars.

Quand peut-on utiliser la promotion aléatoire ?

La promotion aléatoire convient bien pour les évaluations prospectives dans le cadre de programmes universels ou dans lesquels nous ne pouvons pas contrôler qui participe et qui ne participe pas. Elle est particulièrement efficace quand une sorte d'encouragement peut influencer sensiblement l'inscription au programme. La promotion aléatoire n'est pas une bonne solution pour les services qui sont extrêmement populaires, tels que les dons en espèces, dont chacun voudra bénéficier dès qu'il en entendra parler.

Avec cette méthode, nous calculons l'impact moyen de notre programme en nous fondant sur les personnes qui ont participé au programme parce qu'elles y avaient été encouragées. Étant donné que ces participants ne sont qu'un sous-ensemble de la population remplissant les conditions requises, nous avons généralement besoin pour ce type d'évaluation d'échantillons de très grande taille afin d'être sûrs que nos résultats sont *statistiquement significatifs*. Cela alourdit le travail de collecte de données. Si la promotion se fait au niveau communautaire, nous pouvons avoir besoin d'interroger beaucoup plus de personnes dans la communauté que nous ne le ferions avec une simple procédure par tirage au sort. Par conséquent, nos coûts seront probablement plus élevés que les coûts engendrés par d'autres types d'évaluation. D'autres conditions sont présentées dans l'encadré 6.4.

[Définition]

En statistiques, on dit d'un résultat qu'il est **statistiquement significatif** s'il est peu probable qu'il soit apparu par hasard. La signification statistique n'indique rien de l'ampleur de l'effet (signification économique) ; c'est-à-dire que l'impact d'un programme pourrait être statistiquement significatif et cependant très faible.

ENCADRÉ 6.4 Conditions nécessaires pour que la promotion produise des estimations de l'impact valables

Le groupe encouragé et le groupe non encouragé doivent posséder des caractéristiques comparables. On peut y parvenir en assignant aléatoirement des activités de sensibilisation ou de promotion à des personnes, des groupes ou des communautés formant l'échantillon d'évaluation.

La campagne de promotion doit relever sensiblement le taux d'inscription chez les sujets du groupe encouragé par rapport au taux d'inscription chez les sujets du groupe non encouragé. « Sensiblement » désigne un concept relatif fondé sur des besoins en termes de puissance statistique. En général, un programme devrait accroître la participation de 40 % ou plus pour être rentable sur le plan financier. On peut le constater en vérifiant que les taux d'inscription sont supérieurs dans le groupe qui est encouragé à ceux dans le groupe qui ne l'est pas.

Il est important que la promotion elle-même n'influence pas directement les résultats d'intérêt. Si la promotion elle-même change le comportement, il n'est pas possible de déterminer si les changements observés chez les personnes sont dus au programme ou à la promotion. C'est d'autant plus probable si la promotion se fait conjointement avec les programmes de formation. Dans la plupart des cas, il est essentiel de connaître les effets du programme et non de la promotion.

Source : d'après Gertler *et al.* (2011, p. 73).

Avantages

- Dans les campagnes de promotion aléatoire, on ne refuse jamais l'accès d'une personne à un programme ; au contraire, on permet aux personnes de décider par elles-mêmes si elles veulent suivre ou non le programme.

- Ce type d'évaluation produit un groupe témoin de grande qualité présentant, en moyenne, les mêmes caractéristiques que le groupe de traitement, comme toute autre méthode d'aléation décrite plus haut.

Inconvénients

- Cette méthode ne peut être utilisée qu'avec des programmes spécifiques.
- Il est souvent nécessaire de constituer des échantillons de plus grande taille qu'avec les autres méthodes, ce qui augmente les coûts.
- Des techniques statistiques avancées sont nécessaires pour calculer l'impact du programme.
- Les chercheurs doivent être prudents lorsqu'ils interprètent les résultats, car l'estimation de l'impact n'est valable que pour les personnes qui ont participé au programme après y avoir été encouragées ; les résultats ne peuvent pas être généralisés à d'autres groupes de bénéficiaires potentiels.

Un exemple de promotion aléatoire nous est présenté dans l'encadré 6.5.

ENCADRÉ 6.5 Exemple de promotion aléatoire

En Afrique du Sud, une technique de promotion aléatoire a été utilisée pour évaluer l'impact d'un programme d'éducation ludique visant à améliorer les connaissances, les attitudes et les comportements concernant la prise de décisions financières fondées, l'accent étant mis sur la gestion des dettes. Le programme consiste à introduire dans le scénario du soap-opéra *Scandal!*, diffusé pendant plusieurs années, des informations visant à améliorer les compétences financières du public.

Évaluer l'impact d'un soap-opéra sur le comportement et les attitudes est relativement difficile. Premièrement, il est difficile de séparer l'effet du message d'un soap-opéra des autres messages sur des questions similaires que les personnes et les familles peuvent recevoir d'autres sources. Deuxièmement, certains types de personnes peuvent choisir elles-mêmes de voir un certain soap-opéra, ce qui rend plus difficilement observable tout changement d'attitude en résultant. Troisièmement, l'accès à la télévision étant pratiquement universel, il est difficile de constituer un bon groupe témoin formé de personnes qui ne reçoivent pas ces messages visant à améliorer les compétences financières.

Pour surmonter ces difficultés, la méthode de promotion aléatoire ci-après a été élaborée : après identification d'une population d'étude (approximativement 1 000 sujets), environ la moitié de la population a reçu une prime incitative (d'environ 10 dollars) pour regarder *Scandal!* Il s'agissait du groupe de traitement sélectionné au hasard. Pour encourager ces sujets à regarder l'émission, on leur a téléphoné avant la diffusion de trois épisodes sur une période de trois mois pour les avertir qu'une prime leur serait versée pour regarder cette émission. Lors de ces appels, les sujets du groupe de traitement ont été informés des conditions requises pour pouvoir toucher cette prime, et on leur a posé une série de questions pour déterminer leurs connaissances en matière de finances. Après la diffusion de l'épisode, on appelait les personnes et on leur remettait la prime si elles répondaient correctement à plusieurs questions sur le contenu non financier de l'épisode. Au cours de ce même appel, on leur posait également une série de questions sur leurs connaissances financières et leurs attitudes en matière de finances.

La même prime a été versée à l'autre moitié de la population d'étude – le groupe témoin sélectionné au hasard – pour regarder un soap-opéra similaire qui était diffusé à peu près à la même heure et, surtout, qui ne contenait pas d'informations financières à but éducatif. Le mécanisme pour toucher la prime était identique à celui mis en place avec le groupe de traitement. On a posé au groupe témoin les mêmes questions sur leurs connaissances financières qu'au groupe de traitement.

La théorie était que si on parvenait à faire passer le contenu à caractère financier de *Scandal!*, les personnes qui avaient été incitées à regarder le feuilleton répondraient mieux aux questions sur les connaissances financières que le groupe qui était encouragé à regarder l'autre soap-opéra. On a relevé les effets immédiats du projet sur les connaissances et les attitudes au moyen d'une brève enquête après la fin du feuilleton ; les effets à long terme ont été relevés au moyen de multiples enquêtes de suivi.

Source : Banque mondiale (2011).

4e méthode : méthode par discontinuité de la régression

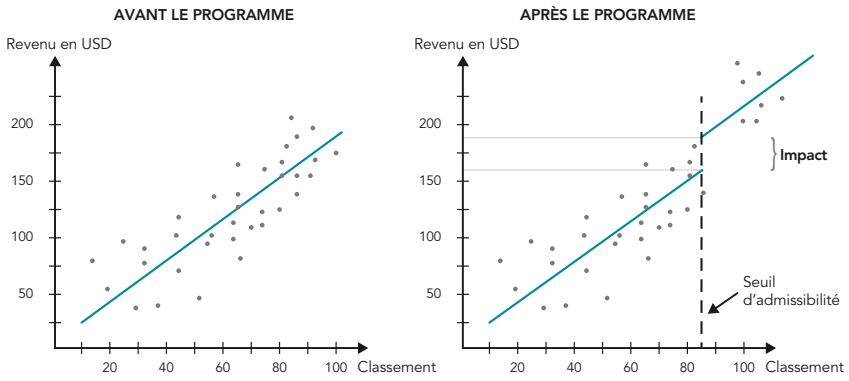
La réalité est que, bien souvent, nous ne sommes pas en mesure de prévoir l'évaluation lors de la conception du programme, et même lorsque nous le sommes, il est parfois impossible d'utiliser toute forme d'aléation pour obtenir un contrefactuel valable. Dans ces cas, nous pouvons peut-être utiliser d'autres règles de ciblage du programme pour obtenir un groupe témoin de qualité. En fait, on procède dans de nombreux programmes à un classement continu des potentiels bénéficiaires, selon les résultats à des épreuves de contrôle, l'évaluation de la solvabilité par score, l'index de pauvreté ou l'âge, et on fixe un seuil pour l'acceptation au programme. Par exemple, les candidats à un concours de plans d'entreprise ou une banque de micro-financement peuvent être notés sur la base d'un ensemble de critères, sur une échelle allant de 1 à 100. Si des jeunes obtiennent une note égale ou supérieure au seuil minimum, disons 85 et plus, ils reçoivent un financement de démarrage. S'ils obtiennent une note inférieure à ce seuil, ils ne sont pas acceptés dans le programme. De tels classements d'admissibilité peuvent être utilisés pour une évaluation d'impact.

Comment procède-t-on ?

Le postulat dans les procédures d'évaluation par discontinuité de la régression (ou indice d'admissibilité) est que les personnes qui obtiennent une note juste au-dessus ou juste en dessous d'un seuil préalablement fixé ne sont pas très différentes les unes des autres, ou du moins la différence peut être continue à travers les notes. Par exemple, les candidats qui obtiennent une note de 86 sont-ils très différents de ceux qui obtiennent une note de 84 ? Probablement que non. Ou les jeunes de 18 ans susceptibles de remplir les conditions requises pour participer à des programmes Travail contre rémunération sont-ils très différents de leurs pairs de 17 ans, qui ne sont peut-être pas admissibles ? Si nous sommes dans une situation où certains de ces jeunes qui bénéficient du programme (grâce à une note juste au-dessus du seuil) et d'autres qui n'en bénéficient pas (à cause d'une note juste en dessous du seuil) ne sont pas fondamentalement différents les uns des autres, alors la comparaison des résultats de ces deux groupes pourrait nous permettre en revanche d'analyser l'impact du programme.

La figure 6.6. illustre ce que nous pourrions constater en analysant l'impact d'une initiative de microcrédit destinée aux jeunes. Le graphique de gauche indique que, au moment de présenter leur demande pour participer au programme, ceux qui ont obtenu de meilleures notes tendaient déjà à avoir des revenus plus élevés. Il doit y avoir de nombreuses raisons à cela : par exemple, les personnes qui avaient un niveau d'éducation quelque peu supérieur gagnent déjà plus, et ce meilleur niveau d'éducation leur a également permis d'obtenir de meilleures notes. Ou les personnes les plus motivées pour créer une entreprise étaient déjà plus entreprenantes, ce que reflètent leurs revenus plus élevés, et leur motivation les a également aidées à convaincre le jury de les soutenir. Beaucoup d'autres explications sont possibles, mais nous n'avons pas forcément besoin de les comprendre pour appliquer cette méthode.

FIGURE 6.6 Graphiques de sélection d'échantillons au moyen de l'approche par discontinuité de la régression



Au lancement du programme, la banque locale de micro-financement a fixé à 85 le seuil pour octroyer un prêt, et tous les candidats ont vu leur demande de prêt acceptée ou refusée en fonction de ce barème. Maintenant, nous aimerions savoir si le programme de microcrédit a eu des effets sur les niveaux de revenu. Comme l'illustre la figure 6.6. (graphique de droite), nous supposons que les personnes qui ont obtenu une note inférieure à 85 ont les mêmes résultats que précédemment, alors que celles qui ont obtenu une note égale ou supérieure à 85 ont globalement augmenté leurs revenus. À partir de cela, il est possible de déterminer l'impact du programme, qui sera représenté par la différence entre les résultats (c'est-à-dire, par la discontinuité de la relation linéaire) proches du seuil.

Quand peut-on utiliser une approche par discontinuité de la régression ?

On peut utiliser l'approche par discontinuité de la régression pour les évaluations à la fois prospectives et rétrospectives. À la différence des techniques d'aléation vues plus haut, on peut également l'utiliser quand le programme est déjà en cours ou achevé. La principale condition est que la participation au programme soit déterminée par une règle de ciblage expressément spécifiée ; en d'autres termes, par une échelle ou un barème continu. Cependant, pour que cette méthode soit efficace, nous devons procéder à de nombreuses observations dans la région juste au-dessus et juste en dessous du seuil fixé afin de disposer d'un nombre suffisant de jeunes que nous pourrons comparer les uns avec les autres. À moins de conduire l'évaluation sans données de base ou de pouvoir tirer parti d'informations relatives à d'autres programmes, nous devrons collecter des données pour utiliser l'approche par discontinuité de la régression, tout comme dans la procédure par tirage au sort, et les coûts seront similaires.

Avantages

- Avec l'approche par discontinuité de la régression, on tire parti des règles de ciblage existantes et on n'a pas besoin d'apporter des changements dans la conception du programme.
- On obtient des estimations sans erreur systématique pour les participants ayant des notes proches du seuil.
- On n'a pas besoin de procéder à aucun type d'aléation, de sorte que cette approche est plus politiquement acceptable que d'autres méthodes.
- L'approche permet de mesurer les effets marginaux potentiels. Par exemple, si on envisage dans le cadre d'un programme de baisser le seuil d'admissibilité de 85 à 75, une approche par discontinuité de la régression peut indiquer quel impact cela aura sur les participants et fournir des informations pour une analyse coûts-bénéfices de la proposition.

Inconvénients

- Avec cette méthode, il faut fixer un seuil précis qui permettra de constituer les groupes.
- Les estimations de l'impact ne sont valables que pour la marge proche du seuil et ne peuvent pas être généralisées aux personnes ayant obtenu des notes loin de ce seuil. La technique ne permet pas d'évaluer l'impact moyen sur les participants au programme.
- Il faut constituer des échantillons d'évaluation de grande taille, car seuls les résultats proches du seuil sont observés et exploités.
- Comme le mentionnent Duflo, Glennerster et Kremer (2006), dans les pays en développement, les règles d'admissibilité sont rarement appliquées strictement tout de suite, et il est fort probable que les groupes ne soient pas distincts, ce qui fait qu'il est difficile d'obtenir des données valables avec cette méthode.

En somme, l'approche par discontinuité de la régression est une bonne solution lorsque l'évaluation est entreprise tardivement ou que l'aléation n'est pas possible. Cependant, on ne peut l'appliquer que dans des circonstances particulières. L'encadré 6.6. présente un exemple de cette approche.

ENCADRÉ 6.6 Exemple d'approche par discontinuité de la régression

Klinger et Schuendeln (2007) adoptent l'approche par discontinuité de la régression pour étudier le rôle de la formation à l'entrepreneuriat dans la création d'entreprises et les résultats de l'activité dans le contexte des concours de plans d'entreprise organisés par l'ONG TechnoServe en Amérique centrale. Le programme offre une formation et des services à la création d'entreprises destinés à aider les participants à élaborer un plan d'activité, et il finance un certain nombre de plans parmi les meilleurs.

Les évaluateurs tirent parti du fait que pour participer au programme, les candidats sont soumis à une présélection au cours de laquelle ils sont notés en fonction de leurs potentielles capacités entrepreneuriales. Le nombre de candidats admis à suivre le programme est fixé avant le lancement du concours. Les candidats sont acceptés à l'atelier si leur note est supérieure au seuil ; si elle est inférieure, ils sont refusés. Cette procédure permet de comparer les bénéficiaires qui ont tout juste réussi avec ceux qui ont raté de peu. Étant donné que les deux groupes ont obtenu des notes proches, juste au-dessus ou juste en dessous du seuil, il est légitime de supposer qu'ils présentent également des caractéristiques non observables semblables, ce qui permet de constituer un contrefactuel de grande qualité.

L'analyse statistique a confirmé que les règles d'admissibilité avaient été respectées – c'est-à-dire que les personnes avaient été sélectionnées en fonction de leur note – et que les résultats des candidats étaient conformes à la note qu'ils avaient obtenue avant l'entrée dans le programme. Au terme du programme, les évaluateurs ont constaté un changement plus prononcé dans les résultats proches du seuil. Avec l'approche par discontinuité de la régression, ils ont été en mesure de montrer que la formation avait augmenté les chances de créer une entreprise d'environ 10 % et les chances de développer son entreprise de plus de 20 %.

5e méthode : double différence

Dans de nombreux programmes, la sélection des zones et des bénéficiaires cibles ne répond pas à des critères clairement établis, ce qui peut entraîner un ciblage très sélectif. Par exemple, nous possédons déjà des informations sur une communauté particulière, nous avons un meilleur accès à certains endroits qu'à d'autres, ou les partenaires existants disposent déjà d'une infrastructure de base dont nous aimerions profiter. S'il n'y a rien de mal à cela, il est cependant plus difficile de développer un contrefactuel de qualité avec des règles de ciblage si subjectives. Néanmoins, nous pouvons avoir une estimation approximative de l'impact d'un programme en utilisant une procédure d'évaluation en double différence.

Comment procède-t-on ?

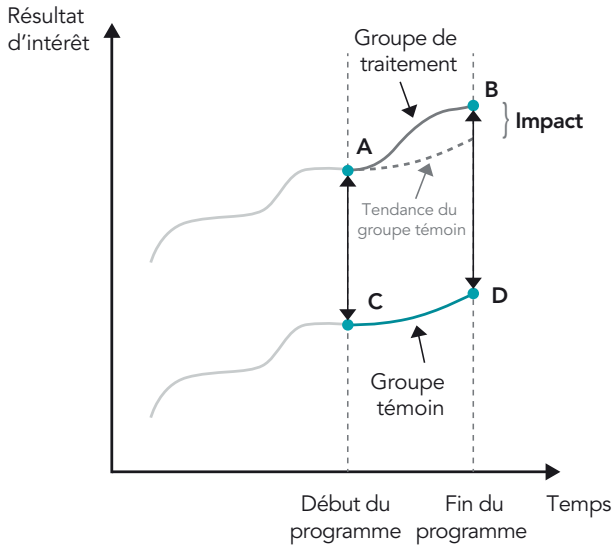
Identifier le groupe témoin. La procédure de la double différence est en fait structurée comme une expérimentation randomisée pré-test/post-test à laquelle il manque sa principale caractéristique, la répartition aléatoire (Trochim 2006). Dans la méthode de la double différence, nous essayons d'identifier un groupe témoin que nous *pensons* être semblable au groupe de traitement préalablement défini. Par exemple, dans les interventions visant à améliorer les moyens de subsistance des jeunes qui ont été menées dans les centres de formation, nous pouvons choisir deux centres ou classes comparables. Dans les programmes communautaires, nous pouvons utiliser deux quartiers ou districts similaires. Dans les deux cas, nous essayons toujours de sélectionner des groupes que nous pensons être aussi semblables que possible, de sorte que nous puissions comparer le groupe

de traitement et le groupe témoin. Cependant, vu que la sélection ne se fait pas de manière aléatoire, nous ne pouvons jamais être sûrs que les deux groupes soient véritablement comparables – rappelons-nous qu'il y a des caractéristiques non observables dont nous ne pouvons pas tenir compte. Aussi parle-t-on d'évaluation par comparaison de groupes non équivalents (Trochim 2006).

Estimer l'impact. Comme nous l'avons vu dans la note 5, la simple comparaison des résultats obtenus par les participants avec ceux des non-participants sélectionnés de manière subjective ne reflète pas l'impact du programme, car les deux groupes sont très probablement différents l'un de l'autre. De la même façon, la comparaison des participants au programme avant et après une intervention pose problème, car beaucoup d'autres facteurs risquent d'influencer les résultats des participants dans le temps. Mais que se passerait-il si nous combinions les deux techniques et comparions les changements survenus avant et après dans les résultats d'un groupe qui a suivi le programme et d'un groupe qui ne l'a pas suivi ?

Imaginons un programme de formation à l'emploi destiné aux jeunes. Pour appliquer la méthode d'évaluation en double différence, nous devons mesurer les résultats (revenu mensuel, par exemple) dans le groupe de traitement et le groupe témoin avant le début du programme (voir figure 6.7., points A et C) et mesurer les résultats des deux groupes au terme du programme (points B et D). Puisque les deux groupes sont probablement différents depuis le début, leurs revenus au départ risquent également d'être différents, mais cela ne signifie pas que la méthode est à rejeter en bloc. La technique de la double différence permet de comparer les différences de résultats entre les deux groupes à la fin de l'intervention (B moins D) avec les différences de résultats entre les deux groupes au début de l'intervention (A moins C). Autrement, nous pourrions comparer les différences de résultats entre les participants (B moins A) avec les différences de résultats entre les non-participants (D moins C). En soustrayant ces différences les unes aux autres, nous avons une vague idée de l'impact du programme ; le résultat montre si le programme de formation a contribué à accroître les revenus des participants, et dans quelle mesure, par rapport à ceux des non-participants.

FIGURE 6.7 Exemple d'analyse en double différence



Source : d'après Gertler *et al.* (2011).

L'hypothèse des « tendances égales ». L'hypothèse sous-jacente de cette méthode est la suivante : bien que les caractéristiques observées et non observées dans le groupe de traitement et dans le groupe témoin soient quelque peu différentes (notamment, niveaux de revenu différents au début), leurs *différences sont constantes au fil du temps*, ou invariables dans le temps. Cela nous permet d'utiliser la tendance du groupe témoin pour estimer ce qui serait arrivé à notre groupe témoin en absence d'intervention.

Une telle hypothèse est-elle réaliste ? De nombreuses caractéristiques observables, telles que l'année de naissance, le sexe, le niveau d'éducation des parents, etc., ne changeront probablement pas au cours de l'évaluation. Cependant, on ne peut pas en dire de même pour plusieurs caractéristiques non observables telles que les traits de personnalité, la motivation personnelle, le goût du risque, etc., qui, comme l'ont montré de nombreuses études, changent au fil du temps, notamment sous l'influence des programmes de développement (voir par exemple Robins et al. 2001 et Roberts, Caspi et Moffitt 2003). Par conséquent, nous ne pouvons jamais être certains que les différences entre les groupes ne changent pas au fil du temps, ce qui pourrait influencer nos estimations de l'impact. Même si les différences observées dans les caractéristiques des participants restent constantes, ces différences pourraient déclencher des effets de propagation au fil du temps. Si les jeunes participants sont, en moyenne, plus motivés que les non-participants, ils pourraient tirer davantage profit du programme et, par suite, obtenir de meilleurs résultats de leur participation que les non-participants. En outre, des facteurs externes peuvent influencer plus ou moins les deux groupes durant la période de mise en œuvre. Ce serait par exemple le cas si la municipalité lançait un nouveau programme dans notre communauté de traitement mais pas dans notre communauté témoin.

[Conseil]

Voici un bon test pour savoir s'il est réaliste de supposer que les tendances sont égales entre les participants et les non-participants : comparez les changements au niveau de leurs résultats avant la mise en œuvre du programme. Si les résultats évoluent en tandem avant le début du programme, nous pouvons être plus sûrs que leurs résultats poursuivront cette tendance au cours du programme. Si par contre les tendances avant le début du programme sont différentes, l'hypothèse des tendances égales peut être erronée. Cependant, connaître la différence entre les tendances nous permettrait au moins de tenir compte de cette différence au moment d'évaluer l'analyse.

Source : d'après Gertler *et al.* (2011).

Quand peut-on utiliser une méthode de la double différence ?

On a avantage à utiliser cette méthode en l'absence d'un mécanisme de ciblage clairement défini (tel que la répartition aléatoire ou les classements d'admissibilité). Vu qu'on suppose que les différences entre les participants et les non-participants sont constantes au fil du temps, on utilisera de préférence cette méthode lorsqu'on dispose de données de qualité obtenues à différents moments avant le début du programme. Il devrait y avoir au moins trois collectes de données : deux avant le traitement et au moins une à la fin. Cela signifie que, à moins de disposer de données sur les participants et les non-participants par d'autres moyens, tels qu'une enquête auprès des ménages déjà en cours, les coûts d'une telle évaluation peuvent être beaucoup plus élevés qu'avec d'autres techniques d'évaluations d'impact.

Avantages

- La méthode de la double différence est un moyen de déterminer les différences entre les participants et les non-participants.
- Elle permet de contrôler individuellement beaucoup d'effets.
- Il n'est pas besoin de procéder à une évaluation prospective si les données nécessaires ont déjà été collectées.
- Elle est utile lorsqu'elle est combinée à d'autres méthodes, car elle accroît leur puissance statistique.

Inconvénients

- Elle produit des résultats moins fiables que les méthodes de répartition aléatoire.
- On ne peut pas l'utiliser seule si on ne se fonde pas sur l'hypothèse que le groupe de traitement et le groupe témoin évoluent de la même façon au fil du temps.

- Pour l'utiliser, il faut procéder à trois collectes de données au moins, alors que deux suffisent avec d'autres méthodes ; cette méthode peut donc coûter beaucoup plus.

Voir encadré 6.7. pour avoir un exemple de cette méthode.

ENCADRÉ 6.7 Exemple de méthode de la double différence

Almeida et Galasso (2008) ont étudié les effets à court terme d'un programme visant à promouvoir le travail indépendant parmi les bénéficiaires d'aides en échange d'un travail social en Argentine. À la suite de la grave crise économique de 2001, le Gouvernement argentin a lancé un vaste programme associant aide sociale et aide à l'emploi, *Jefes*, comprenant une initiative destinée à promouvoir l'activité indépendante appelée *Microemprendimientos Productivos* (Microentreprises productives). Ce programme de microentreprises accordait des dons en espèces pour financer les intrants et le matériel et fournissait un appui technique sous la forme de visites régulières de conseillers.

Pour évaluer les effets du programme en l'absence de données expérimentales, Almeida et Galasso ont utilisé une méthode de la double différence. Ils ont comparé la situation des participants au programme au regard du marché du travail avant et après l'intervention avec celle des non-participants. Pour constituer un groupe témoin valable, ils ont tiré parti de la campagne de promotion du programme, durant laquelle les bénéficiaires de *Jefes* pouvaient s'inscrire pour manifester leur intérêt dans le programme. En limitant le groupe témoin aux personnes qui s'étaient montrées intéressées par l'initiative de microentreprise (mais qui finalement n'y ont pas participé), les auteurs entendaient réduire au minimum les problèmes liés au fait de comparer des personnes intéressées par un travail indépendant (par exemple, en raison de leurs capacités entrepreneuriales ou de leur motivation) avec des personnes qui ne l'étaient pas.

Une enquête initiale sur les ménages a été réalisée auprès de 309 participants et de 244 non-participants en novembre 2004. SIEMPRO, l'office public argentin chargé de suivre et d'évaluer la mise en œuvre des programmes de réduction de la pauvreté, conduisait l'enquête. Les mêmes ménages ont été interrogés une nouvelle fois un an plus tard, fin 2005. Avec seulement deux collectes de données à disposition, les évaluateurs ont dû supposer qu'en absence de programme, les participants et les non-participants auraient suivi des tendances comparables sur le marché du travail (l'hypothèse des « tendances égales »).

Les résultats indiquaient que, vu le taux de participation relativement faible, la relance de l'activité indépendante par l'apport d'un capital de démarrage et l'organisation d'une formation en gestion n'est pas nécessairement une solution attrayante pour tous les bénéficiaires d'aides en échange d'un travail social. En outre, bien que le programme ait contribué à augmenter le nombre d'heures de travail des participants, il n'a pas réussi à accroître de manière significative leurs revenus moyens. Enfin, les personnes concernées n'ont pas toutes bénéficié du programme dans la même mesure, des effets positifs ayant été mesurés uniquement chez les participants ayant un niveau d'éducation plus élevé.

6e méthode : appariement

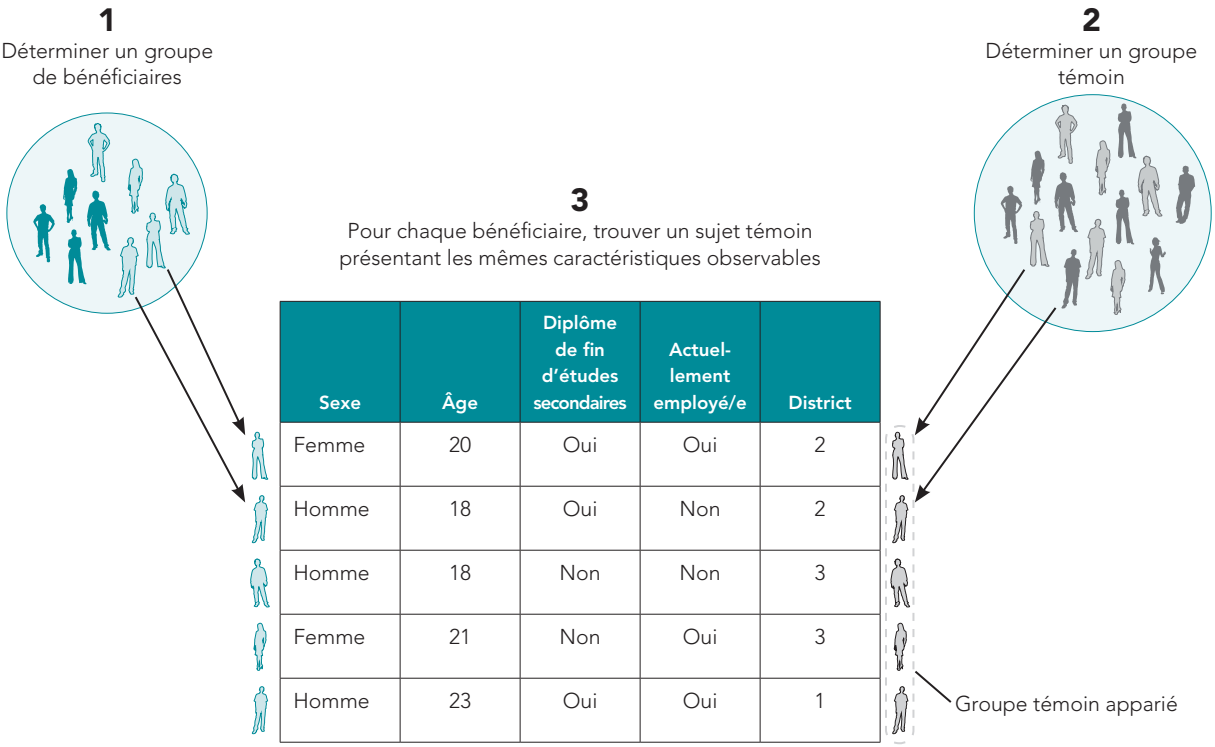
Comme avec la méthode de la double différence, l'appariement est utilisé en l'absence d'autres règles strictes d'assignation à un programme. Par le passé, l'appariement était couramment employé par les spécialistes de l'évaluation des programmes, mais il a peu à peu disparu au profit de méthodes plus fiables, telles que celles décrites plus haut.

Comment procède-t-on ?

Avec la méthode d'appariement, on met ensemble des jeunes participants à un programme et des non-participants sur la base de caractéristiques observables (âge, sexe, niveau d'éducation, situation au regard de l'emploi, résidence officielle et autres facteurs). C'est-à-dire que pour chaque jeune (ou groupe de jeunes) dans la cohorte de traitement, l'appariement permet de constituer une unité témoin artificielle qui présente autant de caractéristiques semblables que possible (voir figure 6.8.). Avec cette technique statistique, on essaie de simuler un groupe témoin qui sinon n'existerait pas.

En fin de compte, les résultats moyens des personnes traitées peuvent être comparés aux résultats du groupe témoin, et les différences reflètent l'impact de l'intervention.

FIGURE 6.8 Appariement exact sur la base de cinq caractéristiques



Sexe	Âge	Diplôme de fin d'études secondaires	Actuellement employé/e	District
Femme	20	Oui	Oui	2
Homme	18	Oui	Non	2
Homme	18	Non	Non	3
Femme	21	Non	Oui	3
Homme	23	Oui	Oui	1

Pour identifier un sujet équivalant à chaque participant au programme, il faut déterminer les caractéristiques qui expliquent qu'une personne décide de s'inscrire au programme. Malheureusement, cela n'est pas aussi facile qu'il y paraît. Comme Gertler *et al.* (2011) le font remarquer, si la liste des caractéristiques pertinentes est courte (comme dans la figure 6.8. ci-dessus), nous trouverons probablement une personne équivalant à chaque jeune du groupe de traitement, mais les appariements ne sont peut-être pas tous précis, et nous risquons de mettre de côté d'autres critères potentiellement importants. En outre, si nous voulons apparier sur la base d'un plus grand nombre de caractéristiques (en ajoutant par exemple le niveau d'éducation des parents, les résultats à un test de contrôle et le niveau de revenu), il peut s'avérer difficile de trouver un équivalent pour chacun des sujets du groupe de traitement, à moins que nous disposions d'un très grand nombre d'observations dans notre base de données de comparaison des jeunes.

Quand peut-on utiliser l'appariement ?

Les techniques d'appariement peuvent être utilisées dans divers cadres, quels que soient la couverture du programme et les critères de ciblage. Dans la pratique, l'appariement est souvent utilisé quand aucune autre procédure d'évaluation n'est applicable, notamment quand l'évaluation commence après le début de la mise en œuvre. Vu la difficulté de tenir compte, avec cette méthode, des caractéristiques non observées, l'appariement est de préférence utilisé en combinaison avec l'une des autres techniques d'évaluation. En outre, pour apparier correctement des unités, et faire en sorte de constituer un groupe témoin qui puisse être apparié, nous avons généralement besoin d'un échantillon de

[Conseil]

Il est possible de surmonter la difficulté de trouver dans le groupe de traitement et le groupe témoin des paires présentant de nombreuses caractéristiques comparables en utilisant une technique appelée l'appariement des coefficients de propension. Au lieu d'apparier des sujets traités et des sujets témoins sur la base de caractéristiques communes pour tous les critères sélectionnés, l'appariement des coefficients de propension permet d'évaluer la probabilité (le coefficient de propension) pour chaque jeune de s'inscrire au programme sur la base de plusieurs caractéristiques observées. Une fois que le coefficient de propension (chiffre entre 0 et 1) a été calculé pour tous les participants et non-participants sur lesquels on dispose de données, les participants sont appariés avec les non-participants qui ont les notes les plus proches. Ces non-participants appariés forment alors le groupe témoin.

grande taille (voir encadré 6.8.). Si les données nécessaires n'ont pas été collectées préalablement par d'autres moyens, ce type d'évaluation risque d'être beaucoup plus coûteux que les autres méthodes décrites plus haut.

ENCADRÉ 6.8 Étapes à suivre pour appliquer une technique d'appariement

1. Identifier les jeunes qui se sont inscrits au programme et ceux qui ne s'y sont pas inscrits.
2. Recueillir des informations détaillées sur les caractéristiques observables (âge, sexe, niveau d'éducation, situation au regard de l'emploi, etc.) des jeunes inscrits et non inscrits au moyen d'une enquête initiale ou en consultant les données existantes.
3. À l'aide d'une technique d'appariement statistique telle que l'appariement des coefficients de propension, apparier chaque participant avec un non-participant semblable.
4. Comparer les résultats des jeunes inscrits avec ceux de leurs équivalents respectifs du groupe témoin. Les différences entre les résultats reflètent l'impact du programme sur chaque bénéficiaire.
5. Calculer l'impact moyen du programme en prenant la moyenne des impacts sur chaque personne.

Avantages

- L'appariement permet de comparer les résultats obtenus par des personnes semblables.
- Il peut être utilisé en combinaison avec d'autres techniques pour valider la qualité du groupe témoin.

Inconvénients

- Comme l'appariement requiert de comparer directement des personnes, il peut être nécessaire de réaliser une enquête sur un échantillon de grande taille afin d'en tirer un groupe témoin approprié.
- L'appariement ne peut être réalisé qu'à partir de caractéristiques observables. Les caractéristiques non observables, ou qu'il est très difficile d'observer, telles que les traits de personnalité, la motivation, le soutien familial, etc., ne peuvent pas être prises en compte dans cette technique. Il faut donc supposer qu'il n'y a pas de différences systémiques dans les caractéristiques non observées entre le groupe de traitement et le groupe témoin, ce qui est rarement plausible. Si cette hypothèse ne tient pas, l'appariement peut conduire à des erreurs systématiques lors de l'évaluation de l'impact du programme.
- Il peut être impossible de trouver une personne équivalente pour chaque sujet du groupe de traitement, ce qui diminue la validité externe de l'évaluation d'impact.

Pour un exemple d'appariement, voir encadré 6.9.

ENCADRÉ 6.9 Exemple d'appariement

Jaramillo et Parodi (2003) ont utilisé l'appariement des coefficients de propension pour évaluer le programme de création d'entreprises par des jeunes, mis en œuvre par l'ONG péruvienne Colectivo Integral de Desarrollo. Pour estimer l'impact du concours de plans d'entreprise et des services de soutien fournis ensuite (formation, suivi des participants et stages), les évaluateurs ont constitué un groupe témoin avec les jeunes qui avaient participé aux activités préparatoires au programme (formation préalable) mais qui n'avaient pas participé au concours de plans d'entreprise ni présenté de propositions gagnantes.

Les évaluateurs ont calculé la probabilité pour chaque personne de participer au programme sur la base de caractéristiques observables telles que l'âge, le sexe, le niveau d'éducation et l'état civil. Chaque bénéficiaire a ensuite été apparié avec un sujet du groupe témoin qui avait un coefficient de propension similaire. La comparaison des résultats (en termes de viabilité de l'entreprise, de nombre d'emplois créés et de revenu) obtenus par les bénéficiaires et leurs pairs respectifs a enfin été utilisée pour évaluer l'impact de l'intervention.

Cependant, étant donné que l'appariement ne pouvait être fondé que sur des caractéristiques observables, le risque était réel que les effets positifs relevés dans l'évaluation soient une surestimation de l'impact réel de l'intervention. En fait, il est probable que les jeunes qui ont réussi le concours de plans d'entreprise soient différents des jeunes formant le groupe témoin, par exemple en en ce qui concerne leur motivation ou leur niveau de compétences, et qu'ils réussissent mieux que leurs pairs dans leur activité entrepreunariale, même sans avoir participé au programme de création d'entreprises.

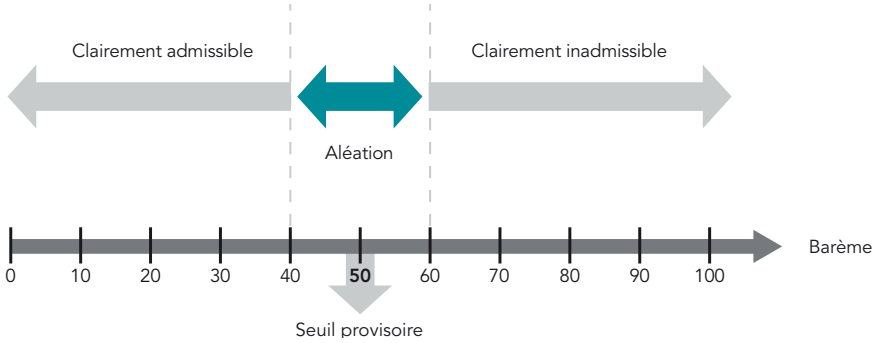
Combinaison de méthodes

Comme nous l'avons vu, certaines méthodes sont plus valables que d'autres pour constituer un contrefactuel. En particulier, il peut être difficile de trouver des groupes témoins de bonne qualité lorsque l'évaluation n'a pas été planifiée dès le début du programme. La combinaison de méthodes peut compenser certaines faiblesses propres à chaque technique utilisée seule et accroître la validité du contrefactuel.

Répartition aléatoire combinée avec l'approche par discontinuité de la régression

Cette technique combine approche par discontinuité de la régression et répartition aléatoire. Si un seuil n'est pas clairement fixé, ou qu'il n'est pas suffisamment justifié, il est possible de procéder à une aléation des sujets ayant des résultats proches du seuil. Dans ce cas, les jeunes qui remplissent nettement les conditions requises bénéficient toujours du programme, tandis que ceux qui ne les remplissent pas clairement n'en bénéficient pas (voir figure 6.9.). Seul un groupe ayant des résultats proches du seuil est sélectionné en vue de l'aléation. Avec cette méthode, certaines personnes qui autrement n'auraient pas pu suivre le programme peuvent maintenant en bénéficier, et vice versa. Comme dans l'approche par discontinuité de la régression ordinaire, les résultats ne sont valables que pour les participants proches du seuil d'admissibilité. Cependant, étant donné l'aléation partielle, nous sommes plus assurés que le groupe de traitement et le groupe témoin présentent les mêmes caractéristiques pertinentes, et nous avons besoin d'un échantillon de plus petite taille pour trouver des résultats statistiquement significatifs. On procède ensuite à l'analyse de la même manière que pour tout dispositif randomisé. Les résultats moyens des sujets du groupe de traitement sont comparés aux résultats moyens des sujets du groupe témoin, et la différence représente l'effet de causalité du programme sur les personnes sélectionnées.

FIGURE 6.9 Barème d'admissibilité (exemple de classement selon le niveau de pauvreté)



Note : Plus la note est faible, plus le niveau de pauvreté est élevé.

Méthode de la double différence ou de l'appariement combinée avec l'aléation

La technique de la double différence suppose que les sujets du groupe de traitement et ceux du groupe témoin sont très similaires, ou du moins que leurs différences sont constantes dans le temps. De la même façon, l'appariement suppose que deux sujets présentent des caractéristiques observables semblables, ce qui justifie qu'on les compare. Avec l'aléation, on ne se fonde sur aucune de ces hypothèses pour évaluer l'impact d'un programme. Cependant, on peut améliorer cette technique en utilisant conjointement l'une ou l'autre des deux premières méthodes. En réduisant au minimum les différences entre les personnes comparées, les méthodes de la double différence ou de l'appariement accroissent la puissance statistique sans qu'il soit nécessaire d'augmenter le nombre de participants. En combinant les méthodes non aléatoires avec les méthodes randomisées, on peut réduire les coûts des enquêtes.

Méthode de la double différence combinée avec l'appariement

S'il est impossible d'utiliser aucun type d'aléation ou d'approche par discontinuité de la régression, l'autre possibilité est de combiner la méthode de la double différence avec la technique de l'appariement, ce qui atténue certaines faiblesses présentées par les deux méthodes quand elles sont utilisées seules. Étant donné que la méthode de la double différence ne peut pas garantir que le groupe de traitement et le groupe témoin sont équivalents, la combinaison de cette méthode avec la technique de l'appariement ou de l'appariement des coefficients de propension peut au moins garantir que les deux groupes sont très similaires en termes de caractéristiques observables.

Pour avoir un aperçu des méthodes d'évaluation ordinaires, voir le tableau 6.1.

[Conseil]

Dans la pratique, le responsable de l'évaluation doit estimer s'il serait utile de combiner des méthodes. Les évaluateurs ne doivent donc pas se soucier des détails concernant la combinaison d'approches mais savoir que ce peut être un moyen d'obtenir des estimations de l'impact plus fiables.

TABLEAU 6.1 Aperçu des techniques d'évaluation d'impact

Méthode	Description	Groupe témoin	Postulats de départ	Données nécessaires	Quand l'utiliser ?
Tirage au sort	Un échantillon de personnes remplissant les conditions requises est réparti de manière aléatoire dans un groupe qui bénéficie de l'intervention et un groupe qui n'en bénéficie pas. L'impact est la différence entre les résultats obtenus par les deux groupes.	Les personnes sélectionnées par tirage au sort.	• L'aléation est efficace et appliquée conformément aux règles. • Les deux groupes sont statistiquement identiques sur des facteurs observés et non observés.	• Données collectées après l'intervention, auprès du groupe de traitement et du groupe témoin. • Données de base souhaitables	• Si l'étude peut être conçue avant le début du programme. • Si les ressources sont limitées et qu'il est important de faire en sorte que des méthodes équitables soient utilisées pour inscrire au programme des personnes dans le besoin. • Si le groupe témoin ne bénéficiera jamais du programme durant toute la durée de l'évaluation.
Procédure par étapes	Les personnes remplissant les conditions requises sont assignées à un traitement par étapes et suivent le programme de manière séquentielle.	Les personnes inscrites sur la liste d'attente.	• L'aléation est efficace et appliquée conformément aux règles. • Les personnes intégrées dans les dernières phases du programme ne changeront pas de comportement de manière significative en attendant de participer au programme.	• Données collectées après l'intervention, auprès du groupe de traitement et du groupe témoin. • Données de base souhaitables	• Si l'étude peut être conçue avant le début du programme. • Si les ressources sont limitées et qu'il est important de faire en sorte que des méthodes équitables soient utilisées pour inscrire au programme des personnes dans le besoin. • Si le programme se déroule sur une longue période.
Promotion aléatoire	Un ensemble de personnes ou de groupes sélectionnés de manière aléatoire sont encouragés à s'inscrire au programme. On mesure l'impact en comparant les résultats moyens obtenus par les personnes qui ont été encouragées à y participer avec les résultats de celles qui ne l'ont pas été.	Les personnes qui n'ont pas bénéficié de la promotion.	• L'aléation est efficace et appliquée conformément aux règles. • Il y aura des différences en termes d'inscription entre les personnes qui bénéficient de la promotion et celles qui n'en bénéficient pas.	• Données collectées après l'intervention, auprès du groupe de traitement et du groupe témoin. • Données de base souhaitables	• Si l'étude peut être conçue avant le début du programme. • Si personne ne peut être exclu du programme. • Si la participation est volontaire et que davantage de personnes participeront au programme si elles sont visées par la promotion du programme.

TABLEAU 6.1 (SUITE) Aperçu des techniques d'évaluation d'impact

Méthode	Description	Groupe témoin	Postulats de départ	Données nécessaires	Quand l'utiliser ?
Discontinuité de la régression	Les personnes sont classées en fonction de critères spécifiques et mesurables. On fixe un seuil pour déterminer qui remplit les conditions requises pour suivre le programme. Les résultats des participants et des non-participants ayant obtenu une note proche du seuil sont ensuite comparés, et les critères d'admissibilité sont pris en compte.	Les personnes ayant obtenu une note proche du seuil mais qui ne remplissaient pas les conditions requises.	• Les personnes ayant obtenu une note juste en dessous ou juste au-dessus du seuil présentent des caractéristiques observées et non observées très similaires.	• Données collectées après l'intervention, auprès des personnes ayant obtenu une note juste en dessous ou juste au-dessus du seuil. • Données de base souhaitables	• S'il n'est pas possible de procéder à une aléation ou que l'évaluation commence après le début du programme. • Si la sélection est fondée sur un classement continu dans lequel un seuil a été fixé.
Double différence	Les résultats des participants au programme et des non-participants sont comparés avant et après l'intervention. Le changement relatif observé dans les résultats reflète l'impact du programme.	Le groupe non équivalent de personnes qui n'ont pas participé au programme mais auprès desquelles des données avaient été collectées.	Au fil du temps, les sujets du groupe de traitement ne changent pas de manière fondamentalement différente par rapport aux sujets du groupe témoin.	• Données de base et de suivi. • Données collectées sur au moins trois périodes (au moins deux collectes doivent être effectuées avant le début du programme).	• Si l'étude commence après le début du programme. • Si l'on peut identifier des non-participants semblables aux participants. • Si l'impartialité de la sélection pour l'entrée dans le programme n'est pas considérée comme un problème. • Plus efficace en combinaison avec d'autres méthodes.
Appariement	Les sujets du groupe de traitement sont appariés avec des non-participants qui présentent des caractéristiques observables similaires.	Appariement exact : pour chaque participant, au moins un non-participant qui présente certaines caractéristiques semblables. Appariement des coefficients de propension : les non-participants présentent un ensemble de caractéristiques qui laisse penser qu'ils auraient autant de chances de participer au programme que les participants.	Les chercheurs peuvent identifier toutes les caractéristiques pertinentes des personnes au moyen d'une enquête.	• Vaste enquête (recensement, enquête démographique et sanitaire, enquête sur la population active, etc.) conduite de préférence en combinaison avec une enquête auprès des ménages dans le cadre d'un programme (prévoir si possible deux observation pour chaque enquête).	• Si l'étude commence après le début du programme. • Si l'on peut identifier des non-participants semblables aux participants. • Si l'impartialité de la sélection pour l'entrée dans le programme n'est pas considérée comme un problème. • Plus efficace en combinaison avec d'autres méthodes.

Points essentiels

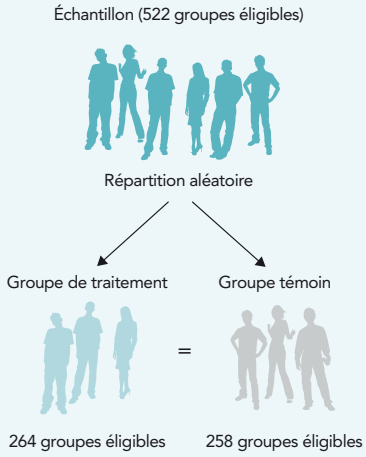
1. Seules certaines méthodes d'évaluation d'impact permettent d'obtenir un contrefactuel fiable et des résultats fondés.
2. La procédure par tirage au sort, la procédure par étapes, la promotion aléatoire et l'approche par discontinuité de la régression permettent d'estimer le contrefactuel au moyen de règles explicites d'assignation au programme. Les méthodes de la double différence et de l'appariement offrent à l'évaluateur des outils supplémentaires – quoique moins précis – pour évaluer l'impact lorsque l'évaluation commence après la mise en œuvre et que les critères d'admissibilité sont moins bien définis.
3. Il n'y a pas de meilleure méthode applicable à tous les programmes. La meilleure méthode dépend du contexte dans lequel le programme est mis en œuvre (à savoir, le moment, la couverture et le ciblage). Par conséquent, les responsables de programmes doivent discuter des contraintes liées au programme avec les spécialistes de l'évaluation, car ces contraintes influenceront l'applicabilité des différentes procédures d'évaluation.
4. Dans la mesure du possible, il est préférable de planifier l'évaluation d'impact avant la mise en œuvre du programme. Les évaluations rétrospectives tendent à être moins fiables, et il est impossible de les conduire si les données nécessaires n'ont pas été collectées par d'autres moyens.
5. Dans certains cas, il n'est pas possible d'appliquer les méthodes décrites ici à cause des contraintes budgétaires, du manque de temps ou pour des raisons politiques.

Étude de cas du NUSAF : le choix d'une procédure par tirage au sort

L'évaluation d'impact du programme « Youth Opportunities Program » du NUSAF, visant à améliorer les perspectives d'emploi des jeunes, a été conçue lors de l'élaboration du programme. Comme le nombre de candidats remplissant les conditions requises pour participer au programme excédait de loin les capacités de financement du programme, le projet d'évaluation d'impact reposait sur la disponibilité d'un grand échantillon de personnes admissibles mais sur des propositions non financées qui avaient été présentées dans le cadre du programme en vue d'un financement. Étant donné cette surabondance de candidatures au programme, la direction du NUSAF et les coordonnateurs du programme ont estimé que la sélection des bénéficiaires au moyen d'un système de tirage au sort était non seulement possible mais que c'était aussi un mécanisme équitable et transparent pour allouer des crédits parmi les jeunes candidats autant qualifiés les uns que les autres.

Les techniciens du NUSAF au niveau des districts ont reçu l'instruction de vérifier que les candidatures satisfassent à un ensemble minimum de critères techniques requis pour l'admissibilité au programme et d'évaluer sur le terrain les projets susceptibles d'être sélectionnés pour être financés. Une liste des projets remplissant les critères requis a été envoyée à la direction du Programme pour que les projets soient présentés à l'équipe chargée de l'évaluation d'impact, qui organisait le tirage au sort destiné à sélectionner les propositions qui seraient financées. Dans chaque district, 30 à 60 % des groupes admissibles ont été sélectionnés pour recevoir un financement, selon les contraintes budgétaires de chaque district.

Une fois que les techniciens au niveau des districts ont eu reçu la liste complète des candidats, on a procédé dans chaque district à la répartition aléatoire des candidats dans le groupe de traitement et le groupe témoin. À l'aide d'un générateur de nombres aléatoires, chaque groupe de candidats a reçu un nombre au hasard. Les groupes ont ensuite été classés par ordre croissant sur la base du nombre qui leur avait été attribué aléatoirement. La somme des coûts du programme a été calculée. En commençant par le premier projet sélectionné aléatoirement, les projets ont reçu un financement jusqu'à ce que les fonds communs de ressources disponibles pour un district donné soient épuisés. Les autres concepteurs de projets n'ont pas reçu de financement et ont été assignés au groupe témoin.



Avec cette procédure, 264 projets au total ont été sélectionnés pour être financés ; ils ont constitué le groupe de traitement. Les 258 autres projets admissibles mais qui n'ont pas été sélectionnés pour être financés ont formé le groupe témoin. Aux fins de l'évaluation d'impact, la constitution d'un groupe témoin équivalent a permis d'estimer le contrefactuel, situation dans laquelle le groupe de traitement aurait été en absence d'intervention.

Source : Blattman, Martinez et Fiala (2011).

À lire

Duflo, E., Glennerster, R. et Kremer, M. 2006. "Using Randomization in Development Economics Research: A Toolkit." Bureau for Research and Economic Analysis of Development (BREAD), Working Paper No. 136. (Pour les lecteurs de niveau avancé.) http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf.

Gertler, P., Martinez, S., Premand, P., Rawlings, L., et Vermeersch, C., 2011, *L'évaluation d'impact en pratique,* Washington, D.C., La Banque mondiale, Les chapitres 4 à 8 sont en rapport avec cette note, http://siteresources.worldbank.org/INTHDOFFICE/Resources/IEP_Ebook_FRENCH.pdf

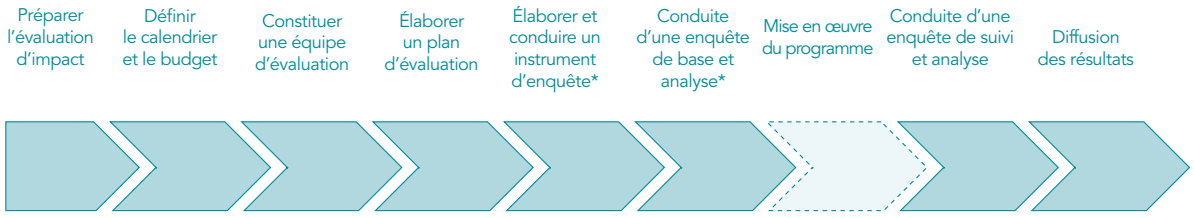
Khandker, S., Koolwal, G., et Samad, H. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. Washington, D.C., La Banque mondiale (Pour les lecteurs de niveau avancé. Les chapitres 3 à 7 sont en rapport avec cette note.) http://www-wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101Official0Use0Only1.pdf

Notes

NOTE 7 : Guide par étape d'une évaluation d'impact

Cette note est un guide par étape visant à évaluer l'incidence des programmes d'intervention sur les moyens de subsistance des jeunes. Les informations contenues dans cette note ne remplacent pas la présence nécessaire d'un spécialiste de l'évaluation d'impact, dans la perspective d'une évaluation correcte. Toutefois, cette note facilitera la planification d'une évaluation d'impact du point de vue du programme, et de la préparation à la diffusion des résultats de l'évaluation (voir figure 7.1.). En outre, elle permettra de clarifier les rôles et responsabilités des acteurs impliqués dans l'évaluation. Nous souhaitons démystifier ce qu'implique une évaluation d'impact, et par là même permettre à chaque organisme ou programme d'envisager d'entreprendre une telle évaluation plus sereinement.

FIGURE 7.1 Étapes à suivre pour mener une évaluation d'impact


Préparer l'évaluation d'impact
Définir le calendrier et le budget
Constituer une équipe d'évaluation
Élaborer un plan d'évaluation
Élaborer et conduire un instrument d'enquête*
Conduite d'une enquête de base et analyse*
Mise en œuvre du programme
Conduite d'une enquête de suivi et analyse
Diffusion des résultats

** Cette étape s'applique uniquement à des méthodes nécessitant une collecte de données par l'organisation.*

Préparer l'évaluation d'impact

Les remarques 2 à 6 de ce guide clarifient les mesures qui devraient être prises avant de commencer une évaluation d'impact. Posez les questions suivantes :

- **Ai-je clairement défini les objectifs de mon programme ?** Les objectifs du programme représentent ce que nous voulons accomplir, et ce que nous attendons de notre intervention. Plus les objectifs sont concrets en termes de population cible, d'ampleur et de calendrier des changements attendus, plus il sera facile de suivre les progrès et de procéder à une évaluation. Par exemple : « En 2015, doubler le revenu de 1 000 dollars E.-U des jeunes de Lima (Pérou) non scolarisés » (voir note 2).
- **Ai-je établi une chaîne de résultats ?** La chaîne de résultats permet aux parties prenantes d'avoir un aperçu fiable, et de savoir comment les ressources et les activités du programme peuvent conduire aux résultats souhaités tout en remplissant les objectifs du programme. Chaque programme devrait être accompagné de sa chaîne de résultats écrite, car c'est la base pour le suivi, mais également pour la définition des questions d'évaluation (voir note 3).
- **Dois-je mettre en place un système de suivi avec des indicateurs et des mécanismes de collecte de données ?** Chaque intervention doit être accompagnée d'un système de suivi préalable. Un tel système nécessite des indicateurs et des techniques de collecte de données, ainsi que l'ensemble des niveaux de la chaîne de résultats clairement définis à dessein d'en suivre la mise en œuvre et les résultats. Sans un bon suivi, les résultats d'une évaluation d'impact risquent d'être d'une utilité limitée, car il est impossible de déterminer si des résultats non satisfaisants sont le fruit d'une mauvaise procédure d'évaluation ou dus à une utilisation faussée de celle-ci (voir note 3).
- **Ai-je formulé par écrit des objectifs d'apprentissage et des questions d'évaluation ?** L'évaluation d'impact devrait être fondée sur nos besoins d'information. Les évaluations d'impact clarifient les relations de cause à effet relatives à certaines questions, c'est-à-dire qu'elles déterminent si les résultats du programme spécifique (généralement un sous-ensemble de ceux définis dans la chaîne des résultats) sont le résultat de l'intervention. Dans la mesure où le type de questions désirées est variable, d'autres outils d'évaluation au-delà de l'évaluation d'impact pour répondre à toutes nos questions sont envisageables (voir note 4).
- **Ai-je identifié un éventail de méthodes d'évaluation d'impact ?** Avant toute chose, une compréhension de base du fonctionnement général d'une évaluation d'impact, et des principales méthodologies pouvant être utilisées est nécessaire. Une fois la connaissance du programme à évaluer acquise, on peut identifier la méthodologie qui convient le mieux au cadre opérationnel existant. Cette connaissance minimale permettra des discussions ultérieures avec les experts en évaluation, et facilitera la planification (voir notes 5 et 6).

Dans les faits, il y a de nombreux malentendus entre les gestionnaires de programme et les experts en évaluation d'impact, car le contexte relatif à l'évaluation n'a pas été clairement défini à l'avance. Une vision précise sur le mode opératoire de l'intervention, et sur les enseignements à retirer d'une évaluation rendra les étapes suivantes plus efficaces, et permettra d'économiser temps et argent.

[Conseil]

Pour voir si votre programme est prêt pour une évaluation d'impact et vous aider à identifier une méthode d'évaluation d'impact appropriée, vous pouvez participer à un atelier d'évaluation d'impact dans lequel il est possible de consulter des experts sur les spécificités de votre programme. Ces ateliers sont offerts par les organisations suivantes :

Le Réseau pour l'emploi des jeunes

http://www.ilo.org/public/french/employment/yen/whatwedo/projects/lead.htm

Laboratoire d'Action contre la Pauvreté Abdul Latif Jameel (J-PAL)

http://www.povertyactionlab.org/course (en anglais)

La Banque mondiale

http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:21754074~menuPK:384336~pagePK:148956~piPK:216618~theSitePK:384329,00.html (en anglais)

Définir le calendrier et le budget

Chronologie

Par définition, le calendrier d'une évaluation d'impact dépend étroitement de celui établi pour le reste du programme. Comme évoqué précédemment (note 6), l'une des questions principales est de savoir s'il est possible de concevoir l'évaluation avant le début de l'intervention, ce qui est à l'évidence toujours mieux. Il est également important de déterminer à quel moment les résultats de l'évaluation sont nécessaires. Si l'obtention de ces résultats est liée à des délais précis, par exemple, pour aiguiller les décisions relatives au programme à mettre en œuvre ou les réformes politiques afférentes, on peut planifier en amont pour s'assurer qu'on dispose du temps nécessaire pour mener l'évaluation d'impact selon la méthode envisagée.

Certaines méthodes sont plus longues à mettre en œuvre que d'autres. Les évaluations prospectives (évaluations prévues à l'avance) ont, comme toutes les évaluations à répartition aléatoire, une échelle de temps plus longue que celles utilisant des techniques rétrospectives, comme la méthode d'appariement. La figure 7.2. recense les principaux facteurs de durée d'une évaluation d'impact. Comme l'illustre le tableau, le calendrier de mise en œuvre de l'évaluation d'impact et le temps nécessaire pour que les effets attendus soient visibles varient d'un programme à l'autre. En règle générale, les évaluations prospectives s'échelonnent habituellement sur une période de douze à dix-huit mois, et les évaluations d'impact rétrospectives sur une période de moins de six mois.

FIGURE 7.2 Calendrier type pour une évaluation d'impact prospective

Tâche	M1	M2	M3	M4	M5	jusqu'à	M16*	M17	M18	M19	M20
PROGRAMME											
Élaborer le programme	■										
Identifier la population éligible		■									
Sélectionner les participants					■						
Mettre en œuvre le programme					■	■	■				
Incorporer les leçons à retenir										■	■
S&E											
Élaborer le système de suivi**	■										
Développer une stratégie d'évaluation d'impact	■	■									
Constituer une équipe pour l'évaluation d'impact	■	■						■			
Élaborer et conduire un instrument d'enquête**		■	■								
Conduire une enquête de base**				■							
Analyser les données de base**					■						
Poursuivre le suivi**					■	■	■				
Conduire une enquête finale**								■			
Analyser les données finales									■	■	
Diffuser les résultats											■

* Dépend du temps nécessaire à la matérialisation des effets

** S'applique uniquement aux évaluations prospectives

Dans les faits, l'échelle de temps plus longue des évaluations prospectives est moins gênante qu'il n'y paraît. Lors de la mise en œuvre de nouveaux programmes, ceux-ci nécessitent généralement plusieurs mois pour être entièrement opérationnels. La préparation d'une évaluation d'impact peut être faite parallèlement à la planification des programmes et des phases d'étude de faisabilité. Celle-ci peut donc aisément être prête lorsque le programme est sur le point d'être lancé. Si un programme a déjà démarré, son organisation en plusieurs phases permet une évaluation d'impact prospective pour la phase suivante du programme.

Budget

Les évaluations d'impact peuvent être onéreuses, ce qui explique la réticence de nombreuses organisations à les financer. Les coûts varient considérablement suivant le pays, la méthodologie utilisée et le type de programme évalué. Le plus souvent, ces évaluations coûtent de 100 000 à plus de 1 000 000 de dollars E.-U. Dans certains cas, notamment lorsque l'ensemble des données est aisément disponible, une évaluation d'impact peut ne coûter que 15 000 dollars E.-U. Si l'évaluation nécessite une collecte de données préalable, la conception et la mise en œuvre d'une évaluation d'impact coûteront probablement plus de 50 000 dollars E.-U.

Facteurs de coût

Dans une évaluation d'impact, les deux sources de dépenses principales sont toujours associées à la durée et au personnel affecté à la collecte des données (voir tableau 7.1.).

Durée. Le temps nécessaire au choix et à la conception d'une méthode d'évaluation appropriée ne doit pas être négligé. Généralement, l'équipe de suivi et d'évaluation peut concevoir l'étude d'impact en collaboration avec un consultant en évaluation. Les honoraires de ce genre de spécialiste varieront, selon expérience, de 200 à 1 000 dollars E.-U par jour, pour un maximum de vingt jours. L'analyse des données nécessite plus de temps, mais elle pourra être conduite par le consultant qui a aidé à concevoir l'évaluation. Par ailleurs, des consultants supplémentaires peuvent être appelés pour se consacrer à des éléments spécifiques de l'évaluation, comme la conception de l'enquête, par exemple. (La prochaine étape, Créer une équipe d'évaluation, fournira plus d'informations sur le rôle et les responsabilités des différents membres de cette équipe.)

La collecte des données. La constitution d'une base de données de référence représente le coût le plus important de l'évaluation d'impact. Faire appel à un institut de sondage est certes plus onéreux qu'une collecte de données réalisée par le personnel du programme, mais elle garantit en principe une meilleure qualité des données. Le coût de référence par personne interviewée pour constituer une base de données de départ dépend de la taille du questionnaire, et de la facilité avec laquelle les personnes interrogées peuvent être trouvées. En cas de questionnaire bref pris en charge par un institut de sondage, avec des personnes aisément identifiées grâce à l'aide de l'équipe du programme, le coût sera de 20 à 40 dollars E.-U par personne interrogée. Dans les lieux où le transport est compliqué et les personnes interrogées difficiles à trouver, le prix peut s'échelonner de 50 à 80 dollars E.-U par personne sondée. Ce prix comprend toutes les étapes de l'enquête, y compris le recrutement et la formation des enquêteurs, l'enquête proprement dite et la présentation des données. Les enquêtes de suivi portent souvent sur des problématiques particulières exigeant le suivi des *participants,* ce qui multiplie leur coût d'environ une fois et demie par rapport au coût de l'enquête de base. Enfin, si l'enquête de suivi n'impose pas de difficultés particulières, et si l'échantillon cible est relativement stable et facile à trouver, celle-ci aura alors un coût inférieur à celui de l'enquête de base.

Les moyens de réduire les coûts peuvent être trouvés en annexe 2.

Évaluation des coûts

Pour la majorité des interventions concernant les moyens de subsistance chez les jeunes, le coût total d'une évaluation d'impact sera vraisemblablement de l'ordre de 150 000 à 500 000 dollars E.-U. C'est une somme relativement importante au vu des nombreux programmes de petite ou moyenne dimension, la question se pose de savoir si ces coûts sont justifiés ou pas.

Répondre à cette question dépend principalement de (1) l'échelle de temps liée au programme envisagé, et (2) les attentes de financement présentes et futures. Par exemple, si l'échelle de temps, même pour un programme relativement modeste avec un budget annuel de 200 000 dollars E.-U est de cinq ans ou plus, ou s'il y a la possibilité d'augmenter le budget jusqu'à 2 000 000 dollars E.-U par

[Ressources en ligne]

Liste des possibilités de financement sélectionnées

http://www.iyfnet.org/gpye-m&e-resource4

L'évaluation d'une formation en matière de finances proposée par FINO en Inde, et mise en œuvre par les banques locales est un exemple d'évaluation pouvant coûter plus que le programme lui-même. Le programme pilote bénéficie à environ 3 000 participants, sa mise en œuvre coûte près de 60 000 dollars E.-U. L'évaluation coûte quant à elle environ 200 000 dollars E.-U. Ce coût a été justifié sur la base de l'évolutivité du projet. Le programme bancaire compte actuellement plus de 25 millions de clients en Inde et augmente de plus de 80 000 personnes par jour. La valeur de l'information délivrée par l'évaluation d'impact ne concerne pas seulement le programme pilote, mais offre aussi cette possibilité à des millions de futurs bénéficiaires.

an, alors les 250 000 dollars E.-U destinés à l'évaluation d'impact, donnant des informations sur la conception du programme, sont un bon investissement. Ne pas réaliser d'évaluation d'impact tout en exécutant un programme marqué par l'inefficacité s'avèrerait *de facto* beaucoup plus coûteux. D'autre part, s'il est établi qu'un programme équivalent n'a qu'une durée de deux ans, alors le coût d'une évaluation d'impact peut s'avérer disproportionné, malgré les connaissances générées par l'étude dont pourraient bénéficier au plus grand nombre de jeunes. Dans ce cas, la décision peut dépendre de la disponibilité de fonds extérieurs au vu du partage des coûts.

TABLEAU 7.1 Exemple de budget d'une évaluation d'impact

	Étape de conception				Étape de référence				Étape de suivi			
	Unité	Coût par unité (USD)	N° d'unités	Coût total	Unité	Coût par unité (USD)	N° d'unités	Coût total	Unité	Coût par unité (USD)	N° d'unités	Coût total
Salaires du personnel												
Directeur du programme	Semaines	2000	2	4000	Semaines	2000	1	2000	Semaines	2000	1	2000
Responsable de suivi et d'évaluation	Semaines	1000	3	3000	Semaines	1000	3	3000	Semaines	1000	3	3000
Honoraires des consultants												
Investigateur principal	Jours	400	10	4000	Jours	400	5	2000	Jours	400	10	4000
Expert enquêteur	Jours	300	5	1500	Jours	300	0	0	Jours	300	5	1500
Coordinateur sur le terrain/Assistant de recherche					Jours	100	80	8000	Jours	100	100	10000
Frais de voyage et de séjour												
Billets d'avion du personnel	Déplacements	3000	2	6000	Déplacements	3000	2	6000	Déplacements	3000	2	6000
Frais d'hébergement et *per diem* du personnel	Jours	150	5	750	Jours	150	5	750	Jours	150	5	750
Billets d'avion du consultant	Déplacements	3000	2	6000	Déplacements	3000	2	6000	Déplacements	3000	2	6000
Frais d'hébergement & *per diem* du consultant	Jours	150	20	3000	Jours	150	20	3000	Jours	150	20	3000
Collecte de données*												
Enquêtes					Jeunes	40	2000	80000	Jeunes	60	2000	12000
Divulgation												
Rapport, impression										5000	1	5000
Séminaire(s)										5000	1	5001
Total coût par étape				**28,250**			**110,750**					**166,250**
Total coût évaluation												**305,250**

* Comprend la formation, l'encadrement, le matériel d'enquête, le personnel sur le terrain (enquêteurs, superviseurs), les transports, etc.

Source : Gertler et al. (2011).

Constituer une équipe d'évaluation

Les évaluations d'impact requièrent un large éventail de compétences, et par conséquence une équipe importante. D'une part, il y a les responsables du programme, qui déterminent s'il est nécessaire de mener une évaluation d'impact, formulent des questions d'évaluation, et supervisent le projet d'évaluation dans sa globalité. D'autre part, il y a les experts en évaluation, généralement des consultants, qui sont responsables des aspects techniques de l'évaluation, notamment du choix méthodologique, de la planification de la collecte des données, puis de leur analyse.

L'équipe de base se compose du directeur de programme ainsi que d'un responsable de suivi et d'évaluation (tous les deux internes à l'équipe), un expert référent en évaluation, en d'autres termes l'investigateur principal, un assistant de recherche qui opère avec l'investigateur principal, et pour les projets d'évaluation impliquant la collecte de nouvelles données, un expert enquêteur, un coordinateur et une équipe sur le terrain (institut de sondage), ainsi que les gestionnaires de données et les agents de saisie. Le tableau 7.2. détaille le rôle et les responsabilités de chaque personne. Selon la taille du programme et de l'évaluation, ainsi que le niveau de compétence de chacun, il est possible que plusieurs de ces tâches soient attribuées à une seule personne.

[Ressources en ligne]

Termes de référence destinés aux membres principaux de l'équipe d'évaluation d'impact

http://www.iyfnet.org/gpye-m&e-resource10

TABLEAU 7.2 Impact d'évaluation – équipe et responsabilités

Qui ?	Tâches principales	Profil/qualifications requises
Directeur de programme	• Définir les objectifs d'apprentissage • Estimer les besoins en ressources • Élaborer le mandat de l'investigateur principal • Engager des consultants pour l'évaluation	• Expérience dans l'élaboration et la mise en œuvre de programmes de développement des moyens de subsistance des jeunes • Expérience dans la gestion d'une équipe • Capable d'élaborer et de gérer des budgets • Capable de travailler dans le respect du programme, et en étroite collaboration avec les équipes d'évaluation
Responsable de suivi et d'évaluation	• Définir le modèle théorique du programme (chaîne de résultats) • Définir des indicateurs et des outils de mesure • Gérer le système de surveillance au démarrage du programme	• Diplôme universitaire de premier ou de deuxième cycle en économie, sciences-politiques ou équivalent • Capable de travailler en étroite collaboration avec le programme et les équipes d'évaluation • Être polyvalent et capable d'assumer des responsabilités en matière de contrôle et d'évaluation
Investigateur principal (université locale ou internationale, groupe de réflexion, conseil spécialisé)	• Sélectionner le projet d'évaluation • Adapter les concepts théoriques à la réalité : budget, temps, données et contraintes politiques • Développer des méthodes mixtes d'approche • Mettre sur pied l'équipe d'évaluation et déterminer les objectifs • Superviser l'équipe • Déterminer les besoins : échantillonnage et moyens • Analyser les données et rédiger un rapport	• Diplôme universitaire en économie, sciences-politiques, ou équivalent • Connaissance du programme ou équivalent • Expérience dans les méthodes de recherche et l'analyse économétrique • Aptitude à travailler efficacement dans des équipes multidisciplinaires • Aptitude à communiquer à l'écrit et à l'oral supérieure à la moyenne
Expert enquêteur (peut être la même personne que l'investigateur principal)	• Conception du questionnaire • Préparer les manuels d'instruction et les livres de codes • Organiser la collecte de données • Mise à l'essai et révision des questionnaires	• Diplôme universitaire en économie, sciences-politiques ou équivalent • Expérience en matière de questionnaires adressés à la jeunesse • Expérience sur le terrain dans le pays ou la région concernée • Aptitude à interagir efficacement avec les autres membres de l'équipe
Coordinateur et équipe de travail sur le terrain	• Collaborer à l'élaboration du questionnaire • Recruter et former les enquêteurs • Établir le calendrier des équipes sur le terrain • Superviser la collecte de données • Nettoyer les données de manière à ce qu'elles puissent être partagées avec le spécialiste d'évaluation	• Statut juridique et carte professionnelle reconnus par le gouvernement du pays où la mission doit être accomplie • Bon réseau d'enquêteurs, de superviseurs et d'agents de saisie expérimentés • Avoir + de 5 années d'expérience effectives dans l'organisation d'enquêtes en rapport avec le programme concerné • De solides aptitudes et une expérience confirmée dans la planification, l'organisation et le suivi logistique d'enquêtes • De solides bases dans la gestion de données et les statistiques • Aptitude à voyager et travailler dans des conditions difficiles
Assistant de recherche	• Analyser les données • Assister l'investigateur principal dans la rédaction des rapports d'évaluation	• Diplôme universitaire de premier ou de deuxième cycle en économie, sciences-politiques ou équivalent
Gestionnaires de données et agents de saisie	• Nettoyer les données pour que l'assistant de recherche et l'investigateur principal puissent les utiliser • Gérer l'équipe affectée à la gestion des données	• Expérience avec les logiciels de gestion de données et dans la direction des équipes concernées.

Après avoir achevé la conception de l'évaluation, après avoir constitué une de base de données de référence et démarré le programme, le directeur du programme et le responsable du suivi et évaluation auront accompli la partie la plus importante de leur mission. Pendant cette période, il est de bon ton de garder l'un deux à mi-temps, sans doute le responsable du suivi et évaluation, pour travailler sur l'évaluation et assurer un suivi adéquat du programme. Lorsqu'il y a des questions importantes relatives à la mise en œuvre du programme, celles-ci devront être documentées, voire dans certains cas, rapportées au reste de l'équipe.

Les experts externes ne seront pas employés au même moment. La première priorité consiste à trouver l'investigateur principal. Celui-ci doit rester en poste pour la totalité de la durée l'évaluation, depuis sa conception jusqu'à la rédaction du rapport final, à dessein d'en assurer la continuité (bien qu'elle ou il ne travaillera probablement pas sur cette évaluation lors de la mise en œuvre du programme). D'autres membres externes de l'équipe peuvent être sélectionnés, si nécessaire, conjointement à l'investigateur principal. Par exemple, l'expert en élaboration d'enquête est habituellement mandaté pour des missions à court terme, et son engagement peut ne durer que quelques semaines suivant l'ampleur de l'évaluation. Les instituts de sondage sont engagés pour mener les enquêtes de référence, et plus tard les enquêtes finales. Idéalement, il faudrait engager le même institut pour les deux types de collecte, bien que cela ne soit pas toujours nécessaire et possible.

Élaborer un plan d'évaluation

Une fois l'investigateur principal en fonction, elle ou il prépare généralement un plan d'évaluation d'impact (appelé également note conceptuelle) en coordination avec les responsables du programme. Ce plan décrira les objectifs, l'élaboration, l'échantillonnage et les stratégies inhérentes à la collecte de données pour l'évaluation. Le plan d'évaluation d'impact (voir schéma type dans l'encadré 7.1.) servira, pour l'essentiel, de base à la méthodologie d'évaluation d'impact, et guidera toutes les étapes du processus de mise en œuvre de l'évaluation.

ENCADRÉ 7.1 Aperçu d'un plan d'évaluation d'impact

1. Introduction
2. Contexte
3. L'intervention
4. La conception de l'évaluation
   - 4.1 Objectif de l'évaluation
   - 4.2 Hypothèses et questions
   - 4.3 Méthodologie d'évaluation
5. La stratégie d'échantillonnage
6. Plan de collecte des données
7. Plan d'analyse des données
   - 7.1 Mesurer l'impact
   - 7.2 Examiner les effets de traitement différentiel
   - 7.3 Mesurer le rendement du programme (analyse coûts-bénéfices)
8. Risques et mesures d'atténuations proposées
9. Audience et diffusion
10. Calendrier et activités
11. Budget
12. Annexes

[Ressources en ligne]

Ressources permettant de trouver les experts en évaluation d'impact

http://www.iyfnet.org/gpye-m&e-resource5

[Conseil]

Le partenariat avec des institutions universitaires est souvent une stratégie efficace pour les ONG et les gouvernements afin de développer leurs capacités à réaliser des évaluations d'impact. Par exemple :

- Save the Children travaille en partenariat avec l'Université des Andes en Colombie pour évaluer l'initiative YouthSave.
- Youth Business International et BRAC sont en partenariat avec la London School of Economics.
- Le Ministère turc du travail a réalisé un partenariat avec l'Université technique du Moyen-Orient pour mener une évaluation sur l'Agence nationale pour l'emploi turque (ISKUR).

L'exemple d'une évaluation d'impact dans le cadre d'un projet de microcrédit pour les jeunes au Yémen montre l'importance de la collaboration, dès le début du programme, entre les personnes chargées de le mettre en œuvre et les évaluateurs, dans le but d'une compréhension mutuelle du contexte opérationnel. Dans cet exemple, les évaluateurs ont défini un essai contrôlé randomisé pour évaluer les répercussions occasionnées par les prêts et autres services financiers adressés aux jeunes sur la création d'emplois, le développement des affaires, etc. Quand le modèle d'évaluation a été présenté au PDG de la banque concernée, celui-ci a trouvé ce modèle inacceptable dans le contexte d'une institution financière récemment fermée, ne pouvant passe permettre d'exclure des clients potentiels pour une évaluation. L'équipe d'évaluation a dû recommencer à zéro et a finalement choisi une méthode d'évaluation plus appropriée dite de promotion aléatoire, qui couvre l'ensemble des clients.

La conception de l'évaluation (point 4) ne doit pas être faite par l'expert en évaluation de manière isolée, mais ce processus devrait au contraire associer étroitement les personnes collaborant à la mise en œuvre du programme pour s'assurer que la méthode d'évaluation correspond bien aux objectifs d'apprentissage et au contexte opérationnel au programme (voir note 6 pour une argumentation détaillée). En outre, malgré un rapprochement probable de l'investigateur principal avec le personnel du programme, et des suggestions visant à définir l'échantillon nécessaire à l'évaluation (point 5) et à la planification de la collecte de données (point 6), il est toujours utile pour l'organisme concerné d'avoir des connaissances de base sur la pertinence de ces aspects pour l'évaluation et le programme lui-même. Aussi, nous explorons ces deux points plus en détail ci-dessous.

Définir l'échantillon pour l'évaluation

Il n'est pas nécessaire de sonder l'ensemble des participants au programme pour évaluer une intervention. Il suffit de choisir un groupe représentatif de personnes, un échantillon assez important pour remplir les objectifs de cette évaluation. Si notre échantillon est représentatif des jeunes qui remplissent toutes les conditions requises, on peut généraliser les résultats de l'évaluation à la population totale éligible. En d'autres termes, on veut que les résultats aient une validité externe, en plus de la validité interne dérivant de la constitution d'un groupe témoin fiable. Pour constituer un échantillon représentatif, nous avons besoin d'une stratégie d'échantillonnage.

Il faut également un échantillon composé d'un nombre assez important d'individus pour obtenir une comparaison fiable des résultats, entre le groupe de traitement et le groupe témoin. Si l'échantillon est trop petit, il peut-être impossible de mettre en évidence les effets statistiquement significatifs du programme, même s'il y en a. Pour mesurer si la taille de l'*échantillon* est assez importante, il faut réaliser des calculs appelés « calculs de puissance ». Ces concepts sont évoqués ci-après.

Créer une stratégie d'échantillonnage

Une stratégie d'échantillonnage comprend trois étapes :

1. **Déterminer la population d'intérêt.** Dans un premier temps, il est important d'avoir une *idée* très *précise* de la population cible et des conditions d'*éligibilité* des participants au programme. Par exemple, l'âge, le sexe, le niveau de revenu, l'emploi et la situation géographique pourraient être déterminants. Les individus non éligibles seront de facto exclus de l'étude.
2. **Constituer une base de sondage.** La base de sondage (ou cadre d'échantillonnage) est la liste la plus exhaustive possible d'unités provenant de la population. Cette base établit un rapport avec l'ensemble de la population d'intérêt, qui est la cible des résultats de l'évaluation. En théorie, pour être totalement représentative, la base de sondage devrait correspondre exactement à la population d'intérêt. Sur le terrain, nous essayons, dans la mesure du possible, d'obtenir une liste des jeunes éligibles exhaustive à partir du recensement de la population, des registres scolaires ou de listes électorales, ou de registres communaux. Cependant, il n'est pas toujours facile d'obtenir une base de sondage qui corresponde à la population éligible toute entière.
3. **Sélectionner le nombre d'unités souhaité dans la base de sondage *via* une des méthodes d'échantillonnage disponibles.** Le prélèvement des échantillons peut être fait à partir de différentes méthodes, mais les plus couramment utilisées sont certaines méthodes d'échantillonnage dites probabilistes. Avec ce type de méthode, les unités de sondage (participants) sont tirées au sort de manière aléatoire, chaque participant ayant la même probabilité d'être sélectionné. Quant aux méthodes non-probabilistes, elles font courir le risque de créer un échantillon peu représentatif de la population éligible dans son ensemble.

Lorsque la liste à disposition n'est pas exhaustive, et qu'il n'est pas possible de savoir avec certitude si la population répertoriée représente l'ensemble de la population d'intérêt, alors il faut éviter

[Définition]

Validité externe : Découle de l'aptitude à généraliser les résultats. En d'autres termes, il faut s'attendre à des résultats similaires si le programme est appliqué à des groupes différents ou plus importants. Pour garantir cela, une stratégie appropriée de l'échantillonnage est nécessaire.

Base de sondage : Consiste à regrouper le plus grand nombre d'individus (unités) parmi la population d'intérêt. De cette liste sera extrait l'échantillon.

Échantillon : L'*échantillon est un sous-ensemble* des éléments d'une *population*. Il est habituellement difficile, voire impossible, d'interroger directement l'ensemble de la population d'intérêt, la solution consiste à recueillir les informations auprès d'un échantillon plus accessible. Si l'échantillon est bien prélevé, on pourra généraliser les résultats de l'évaluation à l'ensemble de la population d'intérêt, via les techniques d'inférence et d'extrapolation.

de généraliser les résultats de l'évaluation au-delà du groupe sondé. En effet, il est toujours tentant de tirer des enseignements généraux au-delà de l'échantillon testé. De nombreuses études tendent à faire cela, mais il faut savoir rester modeste et être prudent dans l'interprétation des résultats. La même prudence est de mise lorsqu'il s'agit de généraliser les conclusions d'une évaluation faite avant l'extension d'un programme. En effet les jeunes concernés par cette extension de programme pourraient être différents de ceux qui étaient impliqués dans l'étude originale.

Les calculs de puissance : « Quelle doit être la taille de mon échantillon ? »

Il est important de connaître la taille idéale de l'échantillon. En d'autre termes, combien d'individus doit-on extraire de la base de sondage ? Si notre échantillon est trop petit, l'analyse statistique peut conclure que le programme sur lequel porte l'évaluation n'a pas de répercussions positives sur ses bénéficiaires supposés, alors qu'il n'en est rien dans les faits. À l'inverse, collecter plus de données qu'il ne faut peut s'avérer très coûteux. Les calculs de puissance permettent de déterminer la bonne taille d'échantillon en indiquant l'échantillon minimum avec lequel il est encore possible d'évaluer l'impact du programme avec un niveau de fiabilité acceptable.

Bien que la taille d'échantillon varie suivant le type d'évaluation, nous estimons pour la nôtre un échantillon de 1 000 à 3 000 jeunes nécessaire pour garantir qu'ils soient assez représentés dans le groupe de traitement et le groupe témoin. Dans certains cas très précis, un échantillon de moins de 1000 jeunes peut suffire. Cependant, il n'est pas recommandé d'avoir moins de 500 participants (250 dans le groupe traité, et 250 pour le groupe témoin). Les professionnels de l'évaluation sont en mesure de calculer la taille de l'échantillon nécessaire pour chaque type d'évaluation.

[Définition]

La puissance représente la probabilité de détecter un impact lorsque celui-ci survient. Il subsiste toujours le risque de ne pas évaluer un impact pourtant bien réel. Toutefois, si ce risque est très faible, on peut considérer le *corpus* de l'étude suffisamment large.

[Ressources en ligne]

Exemple d'estimation de la taille d'échantillon

http://www.iyfnet.org/gpye-m&e-resource6

Planification de la collecte de données

Un plan d'évaluation est nécessaire pour choisir la stratégie de collecte des données de base. La collecte des données peut s'avérer très complexe, aussi vaut-il mieux la confier à des experts extérieurs. Questions clés à se poser :

- Qui du calendrier de collecte des données ?
- Collecter de nouvelles données ?
- Qui va collecter les données ?
- Comment gérer les données ?

Les réponses à ces questions seront évoquées ci après.

[Conseil]

Il peut arriver que des participants se retirent du programme pendant sa mise en œuvre, et abandonnent *de facto* l'évaluation. Aussi est-il prudent de prévoir un échantillon plus important que l'échantillon minimum prévu par les calculs de puissance.

Calendrier de la collecte des données

Ce calendrier est très important et dépend de la nature du programme. Quand une enquête de base est mise en route, elle devrait être achevée avant le début du programme, mais aussi *avant que les participants n'apprennent qu'ils vont bénéficier du programme* pour s'assurer de l'uniformité de leurs réponses *via* le groupe de traitement et le groupe témoin. Ce principe est essentiel, car les jeunes peuvent donner des réponses différentes s'ils savent qu'ils bénéficieront du programme.

Le calendrier de l'enquête de suivi devrait tenir compte des besoins du programme et de ses effets. Si une enquête de suivi est effectuée trop tôt, il n'y aura aucun effet visible, et si elle est faite trop tard le programme ne pourra bénéficier des connaissances acquises.

Dans un programme mis en œuvre en partenariat avec le gouvernement local en Amérique latine, une ONG a connu divers retards concernant la sélection des participants. En effet, un délai assez important s'est écoulé entre la sélection des jeunes concernés, et le début de la formation, ceux-ci ont perdu leur motivation et fini par abandonner le groupe de traitement. En conséquence, le nombre d'unités du groupe de traitement est descendu au-dessous du seuil requis. Dans un tel cas, l'impact devrait être très grand pour avoir une chance d'être quantifiable.

Nouvelles données et données déjà existantes

Il n'est pas toujours nécessaire de collecter de nouvelles données. En effet, dans certains cas, les données requises pour une évaluation existent déjà (voir l'encadré 7.2. qui présente des suggestions pour trouver ces données). Généralement, deux types de données préexistantes doivent être considérées avant de décider d'en collecter de nouvelles.

ENCADRÉ 7.2 Les sources potentielles de données

Données administratives. Ces données sont généralement collectées via un programme de mise en œuvre à des fins de suivi.

Données d'enquêtes sur les ménages. Des enquêtes nationales de ce type sont menées périodiquement dans de nombreux pays en développement. Il s'agit notamment d'enquêtes omnibus, par exemple, les enquêtes sur la mesure du niveau de vie, les enquêtes démographiques et de santé. Ces enquêtes peuvent couvrir un large éventail d'informations sur les caractéristiques du logement, la consommation des ménages, la richesse, le travail individuel, l'éducation, et les indicateurs de santé. D'autres enquêtes, notamment celles concernant les forces de travail, sont plus limitées dans leur portée et ne couvrent parfois que les zones urbaines.

Où regarder :

- Les instituts d'enquête dans les pays respectifs
- Le Réseau international pour les enquêtes auprès des ménages (IHSN) (www.ihsn.org)
- Enquêtes démographiques et de santé (http://www.measuredhs.com/)
- Les enquêtes sur la mesure du niveau de vie (Living Standards Measurement Surveys) (http://iresearch.worldbank.org/lsms/lsmssurveyFinder.htm)

Données du recensement. La plupart des pays organisent tous les dix ans un recensement de leur population ayant pour objectif le dénombrement des logements et de la population. De nombreux pays en profitent pour mener des enquêtes supplémentaires. L'avantage des données du recensement, c'est qu'elles couvrent toute la population. Aussi, existe-t-il des données donnant matière à tout type de traitement et de comparaison. L'inconvénient des données du recensement, c'est que ce dernier est peu fréquent et intègre généralement un nombre restreint d'indicateurs, ce qui limite leur valeur pour une évaluation d'impact.

Où regarder : Réseau international pour les enquêtes auprès des ménages (International Household Survey Network) (www.ihsn.org)

Données d'enquêtes sur les prestations de services. Ces enquêtes collectent des données sur l'efficacité des prestations de services, notamment dans les écoles ou les centres de formation professionnelle. Les ministères et autres entités étatiques, voire les autorités locales peuvent recueillir les informations. Dans de nombreux cas, ces enquêtes sur les prestations de services fournissent des variables de contrôle (par exemple, le rapport enseignant/élèves), tandis que d'autres révèlent des résultats d'intérêt, comme les taux de fréquentation.

Où chercher : Ministères nationaux et représentants locaux

Données d'enquêtes spécialisées. Ces enquêtes ont pour but decollecter des données pour des recherches sur un sujet particulier. De nombreuses enquêtes de ce type reprennent des modules de l'enquête nationale préexistante sur les ménages, et y ajoutent des questions sur des sujets d'intérêt spécifiques. Ces enquêtes spécialisées peuvent avoir une couverture assez limitée, qui, selon les cas, correspondra peu ou pas du tout aux domaines du programme. Néanmoins, si l'équipe d'évaluation peut trouver des données préexistantes issues d'une enquête spécialisée relative à un sujet lié à l'évaluation, ces données peuvent fournir quantité d'indicateurs pertinents.

Où regarder : Les responsables locaux, les donateurs et les ONG dans le domaine d'intérêt.

Source : Banque mondiale (2007a, p. 8-11).

Premièrement, **les données nécessaires peuvent déjà avoir été collectées sous forme de données administratives et de données de suivi et d'évaluation.** En effet, selon les questions auxquelles le programme se propose de répondre, certaines réponses peuvent déjà avoir été recueillies. Par exemple, dans les programmes de subsistance de nombreuses informations relatives au revenu et à l'emploi ont été recueillies au moment du lancement, minimisant ainsi le besoin d'une base de données de référence. Cependant, ces informations ne regardent en principe que les seuls participants au programme. Les données doivent également être collectées au sein du groupe témoin. Pour éviter

d'introduire accidentellement des biais dans l'estimation par la collecte de données incohérentes, il est primordial que le système prévu pour la collecte des données soit le plus cohérent et objectif possible pour les groupes de traitement et témoin. Cette cohérence et cette objectivité sont souvent difficiles à atteindre en cas de collecte de données exclusivement administratives. Dans le cas où un tel système ne fait pas partie du programme, il est préférable d'utiliser une équipe spécifique pour collecter de nouvelles données auprès des groupes de traitement et témoin.

Deuxièmement, **l'institut d'enquête concerné peut avoir déjà collecté des données** auprès de la plupart des personnes participant au programme, et celles composant le groupe témoin. Pour les petits programmes, il est peu probable que suffisamment de personnes impliquées aient déjà pris part à une enquête. Pour les grands programmes, il est *a fortiori* probable qu'au moins certaines l'aient été. Il est également important de comprendre quelles données ont été collectées, et comment celles-ci ont été collectées. Il faut s'assurer que les questions posées portent bien sur le programme dont il est question, et que la taille de l'échantillon est suffisamment importante pour obtenir des conclusions probantes. Il faut également vérifier avec l'institut d'enquête concerné que les données existent et peuvent être utilisés.

Si les informations existantes ne suffisent pas, il faudra se résoudre à collecter de nouvelles données.

Équipe de collecte de données en interne ou en externe

La collecte de données est la partie la plus coûteuse d'une évaluation, et ce pour une bonne raison. En effet, une collecte de données de haute qualité qui soit facilement analysable est la clé d'une évaluation réussie. En d'autres termes, sans données de haute qualité l'évaluation ne vaut plus rien. Lorsqu'il faut décider entre le recours à un institut de sondage ou la collecte interne de données, les responsables du programme doivent choisir la méthode qui s'adapte le mieux à leur budget tout en assurant une qualité et une collecte systématique des données. Certains responsables de ces programmes veulent organiser la collecte de données à l'interne par soucis d'économie. Recourir à une telle solution est viable, à condition qu'il s'agisse d'enquêtes simples et étalées sur de courtes périodes, mais dans le cas contraire, il y a des inconvénients non négligeables, en particulier pour les collectes de données importantes inscrites dans la durée. En raison de la complexité logistique que représente une collecte de données, il n'est habituellement pas conseillé de faire appel pour cela au personnel normalement affecté au programme. Le recours à un institut de sondage est généralement plus coûteux que d'organiser la collecte de données à l'interne, mais cela signifie que les données pourront être collectées plus rapidement et avec moins de travail de la part du personnel affecté au programme. Un institut de sondage garantit également que la collecte de données soit menée par du personnel qualifié (des indications supplémentaires sur l'assurance qualité sont présentes dans les sections Former l'équipe de terrain et Superviser la collecte de données, ci-dessous). De plus, recourir à une entité extérieure assure la neutralité et augmente la crédibilité des résultats de l'évaluation.

Processus de collecte des données et technique

Habituellement, les enquêtes devraient être menées par du personnel qualifié. Les questionnaires auto-administrés ne doivent être utilisés que dans certaines circonstances. Lorsqu'une personne remplit un questionnaire toute seule, elle interprète souvent les questions différemment de ce qui était prévu par les sondeurs. Des enquêteurs formés assurent une plus grande cohérence dans l'interprétation. En outre, il n'est pas rare que les interviewés soient moins doués pour la lecture qu'on ne pourrait s'y attendre, de sorte qu'ils peuvent avoir besoin d'être assistés pour répondre au questionnaire.

Il y a plusieurs manières de collecter et d'enregistrer les réponses d'un questionnaire. Les questionnaires papier sont les plus traditionnels. Les enquêteurs peuvent aussi utiliser des téléphones portables (sur lesquels peut être téléchargé le logiciel relatif à l'enquête), des ordinateurs ou des

[Ressources en ligne]

Protocole pour engager un institut de sondage

http://www.iyfnet.org/gpye-m&e-resource7

[Conseil]

Dans certains cas, un partenariat avec des organisations chargées de mettre en œuvre la collecte de données *via* le personnel des programmes a été conclu. Il n'est pas conseillé d'avoir du personnel dépendant du financement du programme pour procéder à la collecte de données, car il y a un conflit d'intérêt évident en faveur du programme. S'il est décidé d'organiser la collecte de données à l'interne, il est préférable de faire cela avec une équipe distincte, qui se concentre uniquement sur la collecte et ne soit pas directement associée au programme.

[Ressources en ligne]

Outils de collecte de données basés sur les TIC

http://www.iyfnet.org/gpye-m&e-resource2

assistants numériques personnels lorsque ceux-ci sont mis à leur disposition. Il est aussi possible d'enregistrer les réponses des interviewés. Le recours à cette technologie peut nécessiter une formation initiale relativement simple dans la plupart des cas. Ces outils technologiques permettent de réduire le temps nécessaire à chaque entretien et à la saisie des données, ainsi que de diminuer les erreurs dérivées de la saisie et du traitement des données *via* les outils traditionnels. Ces outils peuvent donc faire économiser temps et argent, surtout pour les grandes enquêtes. Toutefois, on doit également considérer s'il est judicieux d'utiliser de tels équipements, parfois coûteux, auprès des ménages démunis et dans les quartiers défavorisés.

Élaborer et conduire un outil d'enquête

Si le plan d'évaluation requiert une nouvelle collecte de données, il est important de choisir le bon outil de collecte. Dans la plupart des cas, ces données seront collectées à travers des sondages, souvent en combinaison avec d'autres méthodes qualitatives, comme les groupes de discussion ou les entretiens avec des informateurs clé.

Dans la mesure où la collecte de données auprès des participants et du groupe témoin est basée sur l'enquête proprement dite, la conception de cette dernière est primordiale. Élaborer des questionnaires peut paraître banal, mais pour une évaluation de haute qualité donnant des résultats fiables, l'élaboration de ces questionnaires est une science, voire un art. Interviewer des adolescents et plus largement des jeunes entraîne des défis supplémentaires par rapport au sondage des adultes. Aussi est-il judicieux de demander l'aide à un consultant expert à ce stade de l'évaluation (voir encadré 7.3.).

ENCADRÉ 7.3 Facteurs influant sur la fiabilité des données au cours d'enquêtes s'adressant aux jeunes

Toute évaluation repose sur l'obtention d'informations fiables. D'après les recherches en la matière, les jeunes répondent généralement de manière fiable aux questionnaires. Malgré cela, ils ont parfois tendance à faire plus de fausses déclarations ou à falsifier plus de réponses que les adultes. Cela peut être dû au facteur suivant :

- **Compréhension.** Les jeunes peuvent avoir moins d'éducation et des capacités cognitives plus limitées que les adultes. Le répondant comprend-il la question ? La question posée utilise-t-elle un langage approprié à l'âge du répondant ? La subtilité de certaines questions peut être difficile à percevoir par un public jeune, même lorsque la question a été posée de manière simple et directe.
- **Rappel.** Quelle probabilité y a-t-il que le répondant se souvienne des événements ou des informations ? Ce souvenir est en partie lié à la période de référence : depuis combien de temps l'événement s'est-il produit ou à quelle fréquence l'événement se produit-il. En général, plus les périodes de rappel sont courtes, plus elles sont précises.
- **Confidentialité.** Le répondant a-t-il une raison de craindre des représailles ou autres conséquences résultant de ses réponses ? Cet entretien est-il réellement privé ? L'intervieweur doit pouvoir convaincre l'interviewé de la confidentialité de l'information.
- **Désirabilité sociale.** Le répondant croit-il que l'enquêteur attend une réponse plutôt qu'une autre ? Une réponse peut-elle perçue comme « correcte » ? Cela concerne en particulier les comportements illégaux, stigmatisés, ou soumis à des restrictions morales. Les études de Brener, Billy et Grady (2003) montrent que les adolescents sont plus à même de déclarer leur consommation d'alcool récente dans des questionnaires auto-administrés qu'ils ne le seraient lorsqu'ils sont interviewés par un enquêteur. En revanche, il n'y a aucune différence dans les réponses des adultes. En outre, de nombreuses études ont démontré que les jeunes sont plus susceptibles que les adultes de donner des réponses incohérentes dans les enquêtes répétées dans le temps.
- **Fatigue.** Les enquêtes s'adressant aux adultes peuvent s'accomplir en plusieurs heures. Dans ces conditions, les jeunes sont susceptibles de perdre patience plus vite que leurs aînés. Par exemple, l'ONG Save the Children a créé l'indice de développement des moyens de subsistance des *jeunes* (*YLDI*). La collecte de données s'articule autour de trois questionnaires auto-administrés adressés aux jeunes de 11 à 24 ans permettant d'obtenir des informations sur leurs biens et compétences. Les questionnaires d'essai ont révélé que les jeunes « s'ennuyaient avec les questionnaires longs et les réponses convenues » (Bertrand *et al.*, p. 5).

Sélection des outils d'enquête, comprenant les questionnaires de base et de fin d'étude du NUSAF.

http://www.iyfnet.org/gpye-m&e-resource11

Note : Le questionnaire du NUSAF est très long, et même s'il se fonde sur une enquête précédente, son pré-test a nécessité l'emploi d'une personne à plein temps pendant quatre semaines. La plupart des questionnaires n'offrent pas autant de questions, mais celui du NUSAF est un bon exemple du type de questions qui peuvent être utilisées dans les programmes de développement des moyens de subsistance des jeunes. Il est également important de tenir compte de la difficulté de mesurer certains paramètres, comme par exemple, les comportements à risque, la santé mentale, l'autonomisation. À chaque enquête son approche, et il est recommandé d'utiliser des outils d'enquête préexistants, de préférence ceux ayant été validés scientifiquement, pour profiter d'une ligne de conduite.

Concevoir et tester l'enquête

Avant de commencer l'enquête sur le terrain, il faut élaborer un questionnaire, selon un processus itératif d'une durée moyenne d'un à deux mois.

Étape 1 : Conception

Le questionnaire est basé sur les résultats et les indicateurs précédemment développés. La langue locale, les dialectes, voire l'argot utilisé par les jeunes, sont des aspects importants dont il faut tenir compte. Pour ce faire, il peut être fait appel à un traducteur. Si des sujets sensibles sont compris dans le questionnaire, notamment des questions relatives à la santé mentale ou à la violence, les questions doivent être formulées avec le plus grand soin, et être en adéquation avec les normes et les coutumes locales. Le premier jet inclut généralement des questions qui seront finalement supprimées ou modifiées.

Étape 2 : Examen interne

Une fois le questionnaire rédigé, les autres membres de l'équipe et les parties prenantes, notamment le directeur de programme, le responsable de suivi et d'évaluation, l'investigateur principal, et l'équipe de terrain doivent l'examiner pour confirmer que celui-ci permettra de recueillir toutes les informations nécessaires.

Étape 3 : Conduite

À ce stade, le projet de questionnaire est testé sur le terrain. L'importance de cette étape est souvent négligée, mais elle est essentielle pour obtenir une évaluation de qualité. La période d'essai sur le terrain est cruciale, et vise à confirmer que la longueur du questionnaire, le formatage et la formulation sont appropriés, tout en s'assurant que l'enquête est susceptible de produire des résultats cohérents et fiables. Le questionnaire devrait être testé sur un échantillon de personnes répondant aux mêmes caractéristiques que celles impliquées dans le programme, mais qui ne seront comprises dans l'échantillon final. Cela garantit que les personnes qui reçoivent le questionnaire final ne soient pas influencées par des questions auxquelles elles auraient déjà répondu. Il est également important de tester en amont les procédures utilisées pour localiser les personnes interrogées, et s'assurer qu'elles puissent être trouvées facilement.

Étape 4: Révision

Le projet de questionnaire est révisé pour répondre aux questions soulevées lors du test sur le terrain. Le cas échéant, les étapes peuvent être répétées jusqu'à ce que tous les problèmes aient été résolus.

Formation des équipes d'enquête de terrain

Lorsque le questionnaire est prêt, les équipes de terrain doivent être formées pour les prendre en charge. L'expert enquêteur ou l'institut de sondage doivent élaborer un manuel que les enquêteurs sur le terrain utiliseront comme outil de formation et guide de référence. Le manuel doit exposer au minimum les objectifs et les procédures de l'enquête, notamment les procédures relatives aux difficultés qu'ils peuvent rencontrer sur le terrain. Chaque question doit être expliquée pour que les enquêteurs puissent comprendre les raisons de la présence de telle ou telle question dans l'enquête. En outre, ce manuel doit fournir aux enquêteurs des instructions précises sur la manière de poser chaque question et d'obtenir des informations utiles. L'investigateur principal et le directeur de programme doivent superviser l'élaboration du manuel. L'encadré 7.4. présente un aperçu de manuel d'enquête.

[Conseil]

Les bonnes pratiques pour enquêter auprès des jeunes sont les suivantes :

- Obtenir le consentement éclairé du jeune et de ses parents (voir ci-dessous la section consacrée à la protection des droits de la personne).
- Utiliser le langage familier ou l'argot en cours, le cas échéant.
- Tenir compte de la durée de l'attention chez les jeunes interviewés ; favoriser les questionnaires courts et intéressants.
- Recourir à des questions d'approfondissement pour augmenter la pertinence des réponses ; se référer à un passé récent pour faciliter les réponses faisant appel à la mémoire.
- À l'instar de l'ensemble des répondants, faire preuve de prudence en matière de calendrier et de formulation des questions sensibles.
- Pour faciliter la localisation postérieure des jeunes, il faut recueillir de nombreuses informations sur la famille, les amis, et le voisinage.
- Si des informations sur les ménages sont utiles, prévoir un module d'enquête distinct destiné aux parents ou tuteurs.

[Ressources en ligne]

Manuels de formation pour la collecte de données

http://www.iyfnet.org/gpye-m&e-resource12

ENCADRÉ 7.4 Aperçu du contenu d'un manuel d'enquête

1. Objectifs de l'enquête
2. Fonctions, rôles et attentes des enquêteurs, des superviseurs et autres personnes liées à l'enquête
3. Procédures de vérification de l'exactitude des données.
4. Détail des procédures d'enquête et d'interview (notamment les procédures pour identifier, localiser et communiquer avec les répondants, ainsi que des informations sur la conduite des enquêteurs, la confidentialité, l'objectivité, le rythme des entretiens, la partialité et la profondeur des questions)
5. Instructions générales sur le remplissage du questionnaire et sur le codage.
6. Explications simples sur chaque question
7. Instructions pour finir et vérifier le questionnaire, et remercier les répondants
8. Instructions pour remplir le rapport de terrain et aviser les responsables concernés des difficultés rencontrées

[Conseil]

Il faut garder à l'esprit les normes culturelles et les coutumes locales lors du recrutement et de l'affectation des enquêteurs. Par exemple, il est généralement de bon ton de recruter du personnel féminin pour interviewer des femmes, en particulier pour les questions délicates. Si les répondants (ou leurs tuteurs) ne se sentent pas à l'aise avec un agent enquêteur, il est peu probable qu'ils se laissent interviewer ; ou s'ils le font, les informations fournies risquent d'être incomplètes, inexactes, et donc peu fiables.

Selon la complexité de l'enquête, la formation des enquêteurs peut prendre de quelques jours à plus d'une semaine. La formation doit obligatoirement débuter par le passage en revue de l'ensemble des questions. Puis, chaque enquêteur doit s'entraîner avec un autre de ses collègues. Les enquêteurs doivent être encouragés à poser des questions durant ce processus pour que chacun comprenne l'ensemble des questions. Ce processus doit se poursuivre jusqu'à ce que tous les enquêteurs se soient parfaitement familiarisés avec la totalité des questions. Après la fin de cette la formation, les enquêteurs doivent s'entraîner *in situ* en interviewant au moins cinq personnes correspondant au profil de l'échantillon d'évaluation.

La formation des enquêteurs est à la fois un processus d'apprentissage et un entretien d'embauche. Il faut convoquer à la formation environ 20 % de personnes en plus par rapport à ce qui est réellement nécessaire. Ce surplus permettra de sélectionner les meilleurs enquêteurs.

Si un institut de sondage est engagé, la formation des enquêteurs lui sera confiée. La présence d'une personne responsable du programme peut s'avérer très utile pour répondre aux questions qui se poseront les tout premiers jours de la formation. C'est aussi la dernière possibilité d'éliminer les erreurs dans le questionnaire.

Protection des droits de la personne

Les recherches impliquant la participation de personnes peuvent parfois constituer un dilemme. Lorsque une recherche vise à répertorier de nouvelles connaissances au bénéfice d'un programme spécifique ou dans une perspective plus large, par exemple, lorsqu'il s'agit d'évaluer l'impact d'une intervention visant au développement des moyens de subsistance des jeunes, il y a le risque de considérer les résultats de l'évaluation comme étant plus important que la protection individuelle des participants. De toute évidence, on ne doit pas utiliser les jeunes dans le seul but d'arriver à ses fins. Il existe des procédures qui permettent d'évaluer la l'aptitude de l'enquête à protéger les jeunes.

La protection des intérêts des participants d'une étude peut être garantie si l'on applique les trois principes suivants (NIH 2008, pp. 17-20) :

- **Le respect des personnes.** Ce principe vise à ce que les futurs participants appréhendent clairement les risques et les avantages liés à une évaluation. En pratique, cela signifie que doit exister un processus visant à assurer que le consentement éclairé et la volonté explicite des

jeunes interviewés sont acquis avant de répondre au questionnaire, dont la nature a clairement été comprise.

- **Bienfaisance.** Ce principe suppose qu'il ne sera non seulement fait aucun tort aux jeunes participants, mais qu'en plus on essaiera de maximiser les effets positifs de l'étude.
- **Justice.** Le principe de justice impose que les individus et les groupes soient traités équitablement en matière de répartition des charges et des effets positifs de l'évaluation.

En vue de garantir les meilleurs standards éthiques en matière d'évaluation, de nombreux chercheurs seront amenés à soumettre leur plan d'évaluation d'impact à un examen auprès d'un comité d'examen institutionnel (Institutional Review Board, IRB, en anglais) dans le pays donateur, le pays hôte, voire les deux. Le recours à ces comités est imposé par la loi à toute personne engageant des recherches, lorsque celles-ci sont soutenues par le gouvernement américain et de nombreux autres gouvernements, ainsi que la plupart des universités à travers le monde. Toutefois, lorsque ces contrôles ne sont pas obligatoires d'un point de vue légal, il est de bon ton d'y avoir recours quand même pour toute personne travaillant sur des études impliquant une participation humaine. Idéalement, un comité d'examen institutionnel devrait examiner toute enquête avant même que celle-ci ne soit testée, mais plus sûrement avant que l'enquête définitive ne soit mise en œuvre au sens large. Les IRB sont présents dans chaque université américaine (la meilleure solution en cas de collaboration avec un chercheur installé aux États-Unis) ou par un comité d'examen local. D'autres institutions, par exemple, les instituts nationaux américains de la santé (NIH) ou l'ONG américaine de lutte contre la pauvreté Innovations for Poverty Action (IPA) pratiquent également des examens d'éthique à la demande. L'encadré 7.5. montre le formulaire-type d'un de ces comités d'éthique.

[Définition]

Un **comité d'examen institutionnel**, aussi connu sous le nom de comité d'éthique (CEI) indépendant, est un comité qui a officiellement été créé pour approuver, surveiller et examiner toute recherche impliquant une participation humaine, dans le but de protéger les droits et le bien-être des personnes.

Le consentement éclairé se réfère à la volonté explicite, de préférence écrite, d'une personne (et, si nécessaire, d'un de ses parents ou de son tuteur) à participer à la recherche. Le consentement éclairé exige une information complète de toutes les caractéristiques de l'enquête susceptible d'influencer la volonté d'un jeune d'y participer.

ENCADRÉ 7.5 Formulaire-type d'un comité d'examen (IRB)

Nom de l'étude : _____

Pays et lieu : _____

Date de test et de fin de l'enquête : _____

Enquêteur(s), nom, poste, département, et institution de tutelle de chacun : _____

I. But / Historique / Pertinence de l'étude et ce qui la rend opportune.

II. Conception de l'étude. Comment les groupes de traitement et témoin sont-ils constitués ? Calendrier du programme. Description de toutes les données devant être collectées.

III. Décrire le profil des participants à l'étude et indiquer s'il y a des groupes vulnérables. Une compensation est-prévue pour les participants ?

IV. Décrire le processus de consentement éclairé.

V. Cette étude pourrait-elle avoir des risques ou des bénéfices ?

VI. Comment la confidentialité peut-elle être garantie ?

VII. Divers : protocole d'entente ou lettre d'appui de la ou des organisation(s) partenaire(s), questionnaire(s), formulaire(s) de consentement, certificat de formation relative aux sujets humains (NIH ou équivalent) pour l'ensemble du personnel de recherche.

Source : Innovations for Poverty Action (2010).

ENCADRÉ 7.6 Conseils sur la procédure d'approbation du IRB

Lorsque votre organisation n'a pas reçu l'approbation d'un comité d'examen institutionnel (IRB)

Presque tous les établissements universitaires abritent ces IRB, ainsi qu'un certain nombre d'organismes donateurs et d'ONG internationales. Si vous travaillez en partenariat avec l'un de ces organismes, on peut vous demander ou vous conseiller de suivre une procédure pour obtenir l'approbation d'un IRB. Si vous travaillez de façon indépendante ou si vous n'avez pas accès au comité d'éthique d'un partenaire, de nombreuses universités et autres institutions fournissent des services d'examen d'éthique. Le Bureau de protection de la recherche humaine (Office for the Human Research Protections), rattaché au Ministère américain de la santé et des services sociaux, dispose d'une base de données consultable de plus de 8 000 IRB dans le monde, de l'Afghanistan au Zimbabwe (voir http://ohrp.cit.nih.gov/search/irbsearch.aspx?styp=bsc). En outre, de nombreuses agences indépendantes fournissent une évaluation éthique moyennant en principe une taxe. Pour plus d'informations, consultez le site de l'Association for the Accreditation of Human Research Protection Programs (AAHRPP) (http://www.aahrpp.org/www.aspx), et celui de l'ONG Consortium of Independent Review Boards (http://www.consortiumofirb.org/), qui assurent la *protection* des personnes en matière de recherche.

Quand il n'y a pas assez de temps pour suivre un processus d'approbation IRB complet

En premier lieu, réévaluer la probabilité d'obtenir un examen de l'étude dans le temps restant à disposition. Votre programme a pour but d'intervenir dans la vie des jeunes et leurs familles, et vous devez protéger les personnes qui fournissent des informations, et veiller à ce qu'aucun tort ne leur soit fait. Cependant, l'approbation par un comité d'éthique peut prendre jusqu'à plusieurs mois, et vous pouvez être pressé de commencer la mise en œuvre de votre évaluation. Après analyse minutieuse, il peut s'avérer qu'il y a impossibilité d'obtenir l'approbation d'un IRB au moment opportun. Dans ce cas, il faut au minimum que tous les membres de l'équipe d'évaluation aient reçu une formation sur la protection des personnes qui participent aux programmes et aux recherches afférentes. Les instituts nationaux américains de la santé (NIH) offrent une formation gratuite en ligne (en anglais et en espagnol). Pour plus d'informations, voir : http://grants.nih.gov/grants/policy/hs/index.htm

Le respect des normes éthiques est primordial dans tous les projets de recherche et d'évaluation. Cependant, des problèmes spécifiques peuvent survenir lorsque ces projets impliquent des jeunes, ces derniers nécessitant une attention particulière (voir tableau 7.3.). Ces problèmes rendent l'approbation de l'évaluation par un comité d'éthique encore plus cruciale que pour d'autres types d'évaluations. Aussi, les chercheurs et les consultants engagés dans l'évaluation doivent-ils recevoir une formation explicite relative au développement de l'enfant et des jeunes avant le début de l'évaluation. En outre, des protocoles clairs doivent être élaborés pour définir quelles données seront collectées et comment elles seront utilisées afin de garantir le respect des normes éthiques les plus pointues, et le plus haut niveau de protection des participants.

TABLEAU 7.3 Aperçu des considérations éthiques à prendre en compte lors d'une étude portant sur les enfants et les jeunes

Problèmes	Pourquoi est-ce important ?	Que faire ?
Informations sur les risques et bénéfices des participants	Les jeunes peuvent avoir une aptitude différente de celle des adultes à évaluer avec précision les bénéfices et les risques associés à la participation à un programme ou une recherche. Généralement, ils ont aussi tendance à prendre plus de risques que les adultes, ce qui les rend plus vulnérables aux conséquences négatives potentielles de leur participation à un programme ou une recherche.	• Anticiper les conséquences possibles pour les enfants et les jeunes concernés. Pas de programme ou de recherche possible tant que les conséquences potentiellement néfastes n'ont été éliminées ou atténuées. • Fournir aux jeunes participants des explications sur l'objectif de la recherche et les procédures proposées dans un langage et sous une forme appropriés à leur âge, leur maturité, leur expérience, et leur état. • Fournir une discussion explicite sur tous les inconvénients et les risques encourus par les jeunes gens qui acceptent de prendre part au programme ou à l'évaluation. • Indiquer clairement qu'il n'y a pas d'obligation de participer à l'enquête pour être admissible au programme. • Ne pas susciter d'attentes inconsidérées concernant les bénéfices ou les récompenses inhérents à la participation. • Le cas échéant, fournir uniquement de modestes récompenses ou des incitations à participer qui soient en adéquation avec le niveau du pays ou de la région concernés.
Consentement	Les parents ou tuteurs des jeunes n'ayant pas atteint l'âge légal de la majorité doivent donner leur consentement avant que les jeunes dont ils dépendent ne soient impliqués dans un programme ou une évaluation. Par ailleurs, obtenir une opinion fiable auprès de certains jeunes peut être difficile, car leur socialisation s'exprime souvent par leurs divergences d'opinion avec les adultes, et peu importe qu'ils soient ou non d'accord.	• Déterminer l'âge légal de la majorité dans le pays concerné, et s'informer localement pour connaître qui est en mesure de donner la permission d'impliquer des jeunes dans un programme et une évaluation (parents, enseignants, autorités locales, dirigeants communautaires, etc.) • Travailler avec des mineurs implique de demander systématiquement le consentement (éclairé) de leurs parents ou tuteurs. • Si l'âge, la maturité et la situation des jeunes participants le permettent, il faut aussi obtenir leur consentement en plus de celui de leurs parents ou tuteurs.
Collecte de données	La collecte d'informations sur des sujets sensibles (usage de stupéfiants, relations sexuelles, activités criminelles, etc.) ou des expériences douloureuses (mauvais traitements, perte des parents, privation, etc.) est plus délicate lorsqu'elle implique des enfants et des jeunes. Leur vulnérabilité émotionnelle et physique doit être protégée.	• Avant d'interroger les jeunes, il faut recueillir autant d'informations que possible via différentes sources indirectes (adultes, dossiers administratifs, etc.) • S'informer localement et concevoir les questionnaires, les directives relatives aux groupes de discussion, et les autres outils en fonction des caractéristiques du groupe cible spécifique (par exemple, s'assurer que les outils d'enquête sont adaptés à l'âge des jeunes concernés et compréhensibles). • Lorsque cela s'avère nécessaire, il ne faut pas perdre de vue que certains sujets peuvent être sensibles. Il faut aussi anticiper et répondre aux préoccupations des parents et des jeunes participants. • Indiquer clairement que les jeunes participants peuvent refuser de répondre à tout ou partie des questions, et que ce refus n'aura aucun effet sur leur admissibilité au programme. Ces avertissements doivent être répétés avant de poser des questions sensibles.
Confidentialité et Protection	La protection de la vie privée est toujours essentielle, mais plus encore lorsqu'il s'agit d'interviewer des jeunes et de sujets sensibles. Compte tenu de l'implication des parents ou des tuteurs durant le processus de consentement, ceux-ci en tant que représentants légaux, auront sans doute des compromis à faire entre la confidentialité et l'obligation éthique de protéger les personnes interrogées. Par exemple, la présence des parents lors de l'entretien peut compromettre la vie privée des jeunes. Parallèlement, si le jeune interviewé risque un préjudice, la responsabilité d'en informer ses tuteurs peut être engagée.	• Il ne faut avoir de cesse de préserver la confidentialité des réponses des parents et des jeunes participants. Cette confidentialité permettra également de renforcer la fiabilité des informations fournies. • Ne jamais divulguer une information sur le répondant sans l'autorisation expresse de ses parents. • Planifier un mode opératoire permettant d'intervenir lorsque le répondant fournit des informations laissant à penser que lui ou d'autres personnes courent un danger (maltraitance, négligence domestique, crimes et violences subis), ou nécessitent une assistance médicale, juridique ou autre. • Au début de chaque entretien, indépendamment des conditions apparentes du répondant, informer tous les participants des ressources auxquelles ils peuvent se référer.

ENCADRÉ 7.7 Exemple pratique : la protection des sujets humains

Pour procéder à l'évaluation du programme de formation professionnelle *Mi Primer Empleo* auprès des jeunes urbains au Honduras, la Banque mondiale a engagé le National Opinion Research Center (NORC) de l'Université de Chicago pour concevoir le questionnaire et gérer le processus de collecte de données. Même si le Honduras n'a pas d'exigences réglementaires en matière de collecte de données sensibles impliquant la participation de sujets humains, les termes de référence pour l'évaluation ont requis l'approbation d'un comité d'éthique américain pour valider la manière dont a été conçue la recherche, le plan de collecte des données et des procédures de sécurité des données correspondant aux normes internationales en la matière. Le NORC a donc soumis tous les protocoles de recherche et les questionnaires au comité d'éthique de l'université à laquelle il est rattaché, au vu d'être approuvés avant le début de l'enquête de terrain.

Compte tenu de la nature de la recherche, les enquêteurs et les superviseurs ont été sélectionnés en fonction de leur expérience concernant l'implication de jeunes gens en matière de recherche. Pendant le processus d'inscription au programme, les candidats ont été informés qu'ils seraient invités à prendre part à une enquête à participation volontaire, et que leur décision de participer à l'enquête n'influencerait en aucun cas leur sélection pour les programmes de formation. Par ailleurs, étant donné que l'âge légal de la majorité est de 18 ans au Honduras, l'équipe de collecte des données a demandé le consentement écrit des répondants âgés de 17 ans ou moins, et le consentement oral ou écrit d'un des parents du mineur, ou d'un tuteur, pour valider l'inscription au programme, ainsi qu'un consentement distinct du répondant mineur et de son tuteur pour participer à l'enquête d'évaluation.

Pour garantir la confidentialité des informations, les renseignements personnels ne figuraient pas sur les formulaires d'entretien, ces derniers ne mentionnant qu'un identifiant numérique. Les données personnelles (noms, adresse, etc.) ont été exclusivement mises à disposition des entités publiques concernées (Ministère du travail et Sécurité sociale) dans le but de contacter les jeunes gens enregistrés, tandis que les données contenus dans les formulaires d'enquête (sans données personnelles) ont été remis exclusivement à la Banque mondiale pour analyse.

Source : NORC (2007).

[Conseil]

Pour s'informer sur les principes directeurs relatifs aux approches éthiques liées à la recherche impliquant des enfants et des jeunes, consulter les ouvrages et liens suivants :

Society for Research in Child Development. 2007. *Ethical Standards for Research with Children*. Voir http://www.srcd.org/index.php?option=com_content&task=view&id=68 (en anglais)

Schenk, K. et Williamson, J. 2005. *Ethical Approaches to Gathering Information from Children and Adolescents in International Settings: Guidelines and Resources*. Washington, DC: Population Council. Voir http://www.popcouncil.org/pdfs/horizons/childrenethics.pdf (en anglais)

Conduite d'une enquête de base et analyse

L'enquête de base correspond aux premières données collectées sur le groupe de traitement et le groupe témoin. Comme déjà évoqué, une enquête de base n'est pas nécessaire pour tous les programmes et méthodes d'évaluation d'impact. Toutefois, la collecte de données de base est fortement recommandée, car elle fournit une première évaluation permettant de vérifier que les principes fondateurs de l'évaluation d'impact qui ont été choisis sont valables en pratique. La collecte de données fournit également des informations utiles sur les caractéristiques des bénéficiaires qui peuvent éclairer le programme.

Il existe une autre bonne raison pour mener une enquête de base : cela peut aider à localiser les participants par la suite. Si une enquête de base a lieu, celle-ci doit comprendre une liste des coordonnées des personnes sondées, mais aussi de leurs amis et de leur famille pour les appeler si nécessaire lors de l'enquête de suivi.

Le calendrier

Les données de base doivent être collectées juste avant de débuter le programme. Si elles étaient collectées après l'initiation du programme, celui-ci pourrait influencer la collecte d'informations. Si l'enquête de base est réalisée trop en avance par rapport au début du programme, les données collectées ne reflèteront pas fidèlement la situation des participants au début de l'intervention.

Si nous faisons une évaluation prospective, le groupe de traitement et le groupe témoin devront être constitués avant que le programme ne commence. La décision d'affectation dans l'un des deux groupes ne devra être communiquée aux participants qu'une fois les données de base collectées.

Supervision de la collecte des données

L'assurance qualité est essentielle pour garantir que les données collectées sont de premier ordre. Premièrement, il est important de mener des tests de validité pour s'assurer que les enquêteurs travaillent en conformité *avec les normes de la profession*, et qu'ils atteignent l'objectif de questionnaires hebdomadaires qui leur a été assigné. Il est d'usage de constituer une équipe indépendante pour auditer 10 à 15 % des questionnaires afin de s'assurer que les personnes interrogées existent, et que les données ont été collectées avec précision. Certains avantages peuvent motiver les enquêteurs malgré un travail difficile. En plus des salaires, les enquêteurs reçoivent souvent un *per diem* leur permettant de couvrir les frais d'hébergement et de nourriture lors de leurs déplacements, ainsi que d'autres types d'incitations.

Deuxièmement, des mesures doivent être prises pour protéger les données collectées. Généralement, les informations disparaissent en cas de perte des questionnaires remplis, de vol ou de dysfonctionnement de l'ordinateur. Pour éviter la perte de données, les questionnaires doivent être récupérés le plus tôt possible auprès des enquêteurs et stockés en lieu sûr. Les données informatiques doivent toujours être sauvegardées.

Enfin, il est important de veiller à la qualité de saisie des données. L'utilisation d'outils électroniques de saisie de données, comme les téléphones portables ou les assistants numériques personnels (PDA) permet d'éviter les erreurs de saisie, comme le peuvent aussi les mesures de contrôle de qualité standard, dont la double saisie de données.

Analyse et rapport

Une fois que les données de base ont été collectées, l'investigateur principal et l'assistant de recherche doivent compléter l'analyse de base et établir un rapport. Comme il n'y a pas encore de résultats à répertorier pour le programme qui a été l'objet de l'évaluation, le rapport de base sera composé de statistiques descriptives. Les valeurs moyennes démographiques du groupe de traitement et du groupe témoin devront être comparées pour s'assurer que les deux groupes sont similaires. Il conviendra également de noter les différences statistiquement significatives entre les deux groupes. Toutes les questions issues de la collecte de données devront également figurer dans le rapport de base (voir encadré 7.6. Aperçu d'un rapport de base).

ENCADRÉ 7.8 Aperçu d'un rapport de base

1. Présentation
   - 1.1 Description du programme et de l'évaluation
   - 1.2 L'équipe de recherche
   - 1.3 Présentation du rapport
2. Contexte
   - 2.1 Situation géographique
   - 2.2 Contexte historique
   - 2.3 Contexte scientifique
   - 2.4 Description du programme et des partenaires sur le terrain
3. Intervention
   - 3.1 Sélection des groupes et des participants
   - 3.2 Description des interventions
   - 3.3 Problèmes liés à la mise en œuvre
4. Conception de l'évaluation d'impact
   - 4.1 Objectifs de l'intervention et résultats hypothétiques
   - 4.2 Conception de la recherche et aléation
   - 4.3 Mesures des résultats
     - 4.3.1 Résultats primaires recherchés
     - 4.3.2 Résultats secondaires recherchés
     - 4.3.3 Résultats indésirables
     - 4.3.4 Autres paramètres d'intérêt
     - 4.3.5 Différences de traitement
   - 4.4 Problèmes rencontrés
   - 4.5 Intervention, organigramme d'évaluation et calendrier
5. Enquête de base d'administration
   - 5.1 Enquêtes individuelles et de groupes
     - 5.1.1 Développement d'une enquête de base et pré-test
     - 5.1.2 Recrutement de l'institut de sondage et formation des agents recenseurs
     - 5.1.3 Mise en œuvre d'une enquête de base
     - 5.1. Problèmes et préoccupations
   - 5.2 Autres enquêtes
6. Analyses de base
   - 6.1 Caractéristiques de base des participants
   - 6.2 Calculs de puissance et tests d'équilibre sur les données de base
   - 6.3 Validité externe
   - 6.4 Qualité des données
7. Conclusions
   - 7.1 Discussions
   - 7.2 Interprétation
   - 7.3 Généralisabilité

Annexes

Source : d'après Bose (2010).

Comme évoqué à la note 6, la validité de chaque méthode d'évaluation d'impact repose sur un certain nombre d'hypothèses. L'analyse de base peut jouer un rôle important dans la vérification de ces hypothèses, et confirmer que la méthode d'évaluation choisie peut être utilisée, ou si l'on rencontre des problèmes, comment les résoudre. L'annexe 3 présente une liste de tests de vérification et de falsification permettant d'estimer si les hypothèses sur lesquelles se fonde notre évaluation sont avérées.

Conduite de l'enquête de suivi et analyse

Quand une méthode d'évaluation repose sur la collecte de nouvelles données, l'enquête de suivi ou l'enquête finale fourniront les données souhaitées de longue date qui vont permettre de déterminer si l'intervention est un succès ou pas. Lorsqu'une évaluation est entièrement basée sur des données existantes, son analyse sera menée au cours de cette étape.

Calendrier

Le directeur de programme et l'investigateur principal déterminent conjointement le calendrier de l'enquête de suivi. Les bénéfices du programme ne seront pas immédiatement observables après l'intervention. Aussi, l'enquête de suivi doit être menée après un délai suffisamment long pour que des effets puissent se matérialiser. Ce temps diffère suivant le type de programme et dépend beaucoup des résultats spécifiques recherchés. Par exemple, les jeunes participant à un programme de formation peuvent subir des désagréments à court terme en matière de revenu par rapport à leurs pairs, car ils ne peuvent travailler pendant le temps dévolu à la formation. Cependant, si cette formation apporte des compétences pertinentes, les jeunes formés pourront obtenir un revenu relativement plus élevé à moyen et à long terme. Le calendrier de l'enquête de suivi sera primordial pour identifier les véritables effets de l'intervention.

Évaluer les effets à court et à long terme peut nécessiter plusieurs enquêtes de suivi, qui majoreront le coût l'évaluation, mais la pertinence des résultats peut augmenter considérablement. Les évaluations d'impact qui suivent les groupes de traitement et les groupes témoin sur plusieurs années sont relativement rares, et leurs résultats sont d'autant plus recherchés et appréciés. La conduite de plusieurs enquêtes de suivi permettra également d'analyser l'évolution des résultats du programme au fil du temps. Cependant, si la mise en œuvre du programme est retardée, il se peut qu'entre la fin du programme et celle du cycle budgétaire ou de subventions, il n'y ait pas assez de temps disponible pour mener une enquête de suivi qui permettrait d'évaluer les résultats à long terme. Lorsque surviennent des modifications du calendrier de mise en œuvre, il est donc important de réaligner le cycle d'évaluation.

Suivi

Il existe une différence majeure entre l'enquête de base et l'enquête finale concernant la question du suivi des répondants. Si les jeunes initialement interrogés ne sont pas localisés lors du suivi, l'analyse pourra être biaisée et altérer la pertinence des résultats. Par exemple, si les participants les moins performants manquent à l'appel, l'évaluation surestimera probablement l'impact du programme. Il se peut aussi que les jeunes les plus capables abandonnent. Comme il n'est pas possible de savoir avec certitude si l'attrition des jeunes conduira à sous-estimer ou surestimer l'impact du programme, minimiser cette attrition est essentiel pour réaliser une bonne évaluation. Dans la mesure où il est presque impossible de réinterroger l'intégralité des jeunes sollicités lors de la phase initiale de l'enquête, il faudra faire tous les efforts possibles pour en retrouver un maximum. On considère un taux d'attrition de 5 à 15 % comme acceptable, ce qui signifie qu'au moins 85 % des jeunes doivent être retrouvés, qu'ils fassent partie du groupe de traitement ou du groupe témoin.

[Conseil]

Pour s'assurer que les résultats de l'évaluation finale sont fiables à long terme, il est bon d'inclure des experts externes dans le processus d'examen de l'évaluation de base et le rapport final. Par ailleurs, en diffusant le rapport de base, l'équipe chargée du programme et de l'évaluation peut susciter un intérêt public pour la recherche en cours, et renforcer l'appropriation et le dialogue entre les parties prenantes internes et externes.

[Conseil]

Il est souvent possible d'identifier des indicateurs intermédiaires compatibles avec les résultats anticipés à long terme. Par exemple, l'impact de la formation à l'entrepreneuriat et des programmes de promotion sur la probabilité de création d'une entreprise pourrait ne pas se concrétiser pendant un certain nombre d'années (les élèves quittent l'école, obtiennent un emploi en vue d'acquérir une expérience appropriée, pour finalement envisager de créer leur propre entreprise.) En analysant les indicateurs à court et moyen terme, notamment les compétences en affaires, la création d'entreprise comme choix de carrière et les mesures concrètes prises pour créer une entreprise, il est possible d'obtenir des résultats d'impact intermédiaires sans avoir à attendre plusieurs années.

[Définition]

L'**attrition** se réfère à la perte de participants ou de répondants au sondage. Cette perte représente un problème pour l'évaluation, car les unités qui ont quitté l'échantillon sont très probablement différentes de celles qui sont destinées à les remplacer, ce qui fausse les résultats. L'attrition se produit pour toutes sortes de raisons : perte d'intérêt dans le programme, migration ou tout simplement à cause de la réticence à participer à l'enquête.

Au Moyen-Orient, l'ONG The Syria Trust for Developpement a fourni des cartes téléphoniques prépayées destinées à motiver les jeunes à participer à une enquête. Pour réduire les coûts, l'ONG a demandé aux opérateurs de téléphonie mobile du pays de fournir ces cartes sous forme de dons en nature. Les compagnies de téléphonie mobile ont fourni 10 000 cartes de 2 dollars E.-U chacune, pour un total 20 000 dollars E.-U). Pour les compagnies de téléphone, cette opération a été une bonne publicité à moindre coût.

Le NUSAF a engagé une société pour mener une enquête de suivi des répondants d'une durée de 10 minutes, un an après l'enquête de base et un an avant l'enquête finale. Le questionnaire a été adressé aux répondants pouvant être facilement localisés grâce aux informations de contact mises à jour. Pour ceux qui n'ont pas pu être localisées, les informations ont été recueillies auprès d'amis ou de la famille. Les informations ainsi prélevées ont été conservées pour aider les enquêteurs à retrouver les répondants lors de l'enquête finale, mais aussi pour qu'ils puissent savoir s'il sera facile ou pas de contacter ces personnes.

[Conseil]

Moyens supplémentaires pour faciliter le suivi :

- Demander de l'aide et conseils aux dirigeants, aux fonctionnaires et aux habitants des pays et régions. Les autochtones connaissent sans doute la meilleure façon de trouver quelqu'un.
- Impliquer des enquêteurs du pays ou de la région concernée, car ils sont familiers des lieux et des coutumes locales.
- Si les participants ne sont toujours pas localisables, il faut sélectionner un échantillon aléatoire des individus non localisés pour mener des recherches très intensives à leur égard. Ces unités, une fois retrouvées, représenteront de manière adéquate celles qui ne l'ont pas été.

Le suivi des personnes, particulièrement chez les jeunes très mobiles, peut s'avérer difficile. Voici trois méthodes courantes pour réduire l'attrition :

- **Recueillir des informations de contact fiables au cours de l'enquête de base.** L'enquête de base doit répertorier différents types de coordonnées (adresse postale, adresse électronique, numéro de téléphone, etc.), notamment celles du répondant, mais aussi de ses amis et de sa famille, dans le but d'améliorer la localisation des jeunes pour l'enquête de suivi. L'utilisation de médias sociaux comme Facebook peut également permettre de garder le contact avec les jeunes.
- **Inciter les jeunes du groupe de traitement et du groupe témoin à participer aux futures enquêtes.** Les incitations peuvent se présenter sous la forme de petites compensations financières pour le temps consacré à l'enquête, ou de tirages au sort permettant de remporter des espèces ou divers autres prix. Ces mesures incitatives peuvent être communiquées aux jeunes en amont, lors de l'enquête de base, par exemple, ou par l'intermédiaire de publicités à la radio ou dans les journaux.
- **Utiliser une enquête de suivi.** Dans le cas des évaluations pour lesquelles une longue période sépare l'enquête de base de l'enquête finale (deux ans ou plus), mais surtout pour celles qui pas recours à l'enquête de base, une brève et rapide enquête de suivi peut être mise en œuvre pour évaluer une probable attrition et recueillir des informations supplémentaires. Si le budget du programme est restreint, il est possible de réaliser des enquêtes de suivi par téléphone, pour mettre à jour les informations de contact des répondants, en limitant les visites personnelles aux jeunes qui n'ont pu être joints par téléphone.

Conception et collecte de données de l'enquête de suivi

Il est probable que les équipes du programme et de l'évaluation veuillent ajouter quelques questions supplémentaires par rapport à l'enquête initiale (voir encadré 7.7.). Il peut s'agir de questions relatives à l'assiduité, à l'abandon et aux raisons poussant les jeunes à suivre le programme ou à abandonner. Ces informations peuvent être servir à estimer la mesure dans laquelle les participants ont suivi le programme. Ces nouvelles questions devront être testées et révisées si nécessaire. Généralement, il est préférable de garder les questions de suivi dans un ordre semblable à celui des questions de base pour pouvoir les comparer. Sauf problème majeur avec une question de l'enquête d'origine, il est préférable de conserver le même libellé dans les enquêtes de suivi. Le manuel d'enquête devra également être mis à jour pour refléter tout changement par rapport aux questions d'origine. Il devrait comprendre en particulier les protocoles spécifiques destinés aux participants de l'enquête de suivi.

ENCADRÉ 7.9 Questions courantes destinées à être ajoutées dans une enquête de suivi

- Raisons de la non-participation ou de l'abandon
- Fréquence de présence des participants ou montant des prestations reçues
- Satisfaction
- Note de qualité du programme
- Auto-évaluation des résultats du programme

Enfin, les enquêteurs auront besoin du même niveau de formation et de supervision comme avec l'enquête de base afin d'assurer la meilleure qualité de la collecte des données. Dans la mesure du possible, il conviendrait de choisir les meilleurs enquêteurs du personnel de base pour mener

l'enquête de suivi. Les intervieweurs avec des taux d'erreur élevés ou ceux qui étaient moins fiables devraient être remplacés ou devraient recevoir une formation supplémentaire.

Analyse finale et rapport d'évaluation

Une fois que les données de suivi ont été collectées, le rapport final d'évaluation d'impact peut être établi. Ce document représente le produit principal de l'évaluation. Le rapport final reproduira dans les grandes lignes les informations présentes dans l'enquête de base, en y ajoutant des informations détaillées sur l'administration de l'enquête de fin d'étude et l'analyse de données finale.

L'analyse sera fondée sur les résultats et les variables identifiées précédemment. Dans certains cas, rares du reste, l'analyse peut être faite par simple comparaison des valeurs moyennes entre le groupe de traitement et le groupe témoin (c'est le cas généralement des modèles de loterie). Cependant, dans la pratique, on utilisera souvent une certaine forme d'analyse de régression pour plusieurs variables clés, qui autrement pourraient biaiser les résultats.

L'encadré 7.10. présente un aperçu des sections d'un rapport d'évaluation pouvant être ajoutées à l'analyse de base. Ces informations sont importantes pour s'assurer qu'une personne non impliquée dans l'évaluation puisse interpréter correctement les résultats.

ENCADRÉ 7.10 Exemple d'ajouts au rapport initial après l'enquête finale

7. Administration de l'enquête finale
   - 7.1 Enquête finale individuelle et de groupe
     - 7.1.1 Développement et pré-test de l'enquête finale
     - 7.1.2 Institut de sondage / Recrutement et formation des enquêteurs
     - 7.1.3 Mobilisation et suivi des protocoles
     - 7.1.4 Mise en œuvre de l'enquête finale
   - 7.2 Protocoles qualitatifs
   - 7.3 Problèmes et retards
   - 7.4 Problèmes de qualité des données
8. Analyse de données
   - 8.1 Méthodes statistiques utilisées
   - 8.2 Niveau d'analyse
   - 8.3 Sommaire des résultats
   - 8.4 Analyses auxiliaires
9. Conclusion
   - 9.1 Discussion
   - 9.2 Interprétation
   - 9.3 Généralisabilité
   - 9.4 Orientations pour les futures recherches

Annexes

Source : d'après Bose (2010).

Comprendre l'hétérogénéité

Tous les bénéficiaires du programme n'en profiteront pas de la même manière Par conséquent, la compréhension de la variabilité des impacts du programme est un élément important de l'évaluation. Par exemple, de nombreux responsables de programmes cherchent à savoir qui en bénéficie le plus : les garçons ou les filles ; les enfants ou les adolescents ; les jeunes les plus instruits et expérimentés,

[Définition]

En matière de statistiques, l'analyse de régression comprend toutes les techniques de modélisation et d'analyse de plusieurs variables. Dans l'évaluation d'impact, l'analyse de régression permet de comprendre comment la valeur type de l'indicateur de résultat change lorsque l'assignation au groupe de traitement ou au groupe témoin varie, tandis que les caractéristiques des bénéficiaires restent constantes.

[Ressources en ligne]

Les rapports d'évaluations d'impact

http://www.iyfnet.org/gpye-m&e-resource13

ou les autres. En plus de l'âge, du sexe ou du niveau d'instruction, on peut aussi chercher à évaluer si les résultats diffèrent selon le degré de richesse des participants (actifs), leur capital social (accès aux réseaux), ou leurs traits psychologiques (propension à l'optimisme, au risque, etc.) Appréhender quels participants ont profité le plus, et lesquels ont profité le moins du programme peut aider ses responsables à mieux paramétrer ou cibler l'intervention. Pour plus d'informations sur la mesure d'hétérogénéité, voir Mesure de la diversité des impacts effets à la note 8.)

Par exemple, si l'évaluation révèle qu'un programme de formation sur les moyens de subsistance a eu un grand impact chez les hommes, les futures répétitions du programme pourront être axées davantage sur les hommes pour accroître le rendement global du programme. Selon les priorités, on peut explorer diverses pistes pour que plus de femmes soient impliquées, et qu'elles puissent à leur tour bénéficier du programme.

Toutefois, comme spécifié dans l'encadré 7.11, l'hétérogénéité d'intérêt doit être indiquée préalablement à toute analyse, l'ensemble des résultats doit être rapporté, et pas seulement ceux jugés statistiquement significatifs. Il faut éviter l'exploration de données (data mining), qui s'avère être un problème très important par rapport à l'analyse d'hétérogénéité.

[Définition]

L'hétérogénéité d'impact fait référence à des différences d'impact selon le type de bénéficiaire, autrement dit, comment différents sous-groupes bénéficient d'une intervention à des niveaux différents..

Bruhn and Zia (2011) *ont étudié l'impact d'un programme d'entrepreneuriat et d'éducation financière, sur les résultats d'entreprises gérés par de jeunes entrepreneurs dans l'économie d'après-guerre d'un pays émergent : la Bosnie-Herzégovine. Bien qu'on n'ait pas trouvé d'effets de traitement moyen significatifs du programme de formation sur les performances de l'entreprise, des niveaux élevés d'hétérogénéité ont été relevés parmi les participants. Les jeunes entrepreneurs dont l'éducation financière était relativement élevée avant de prendre part au programme ont amélioré leurs ventes grâce au programme de formation. Les effets sur les bénéfices ont également été positifs pour ce sous-groupe. Ces résultats suggèrent que la formation ne doit pas être la seule intervention visant à soutenir les jeunes entrepreneurs, et que si cette formation s'est révélée pertinente pour un groupe spécifique de jeunes gens, elle ne l'a pas été pour tous.*

ENCADRÉ 7.11 Le data mining (DM)

Le data mining pose un problème majeur en matière de statistiques. C'est particulièrement le cas pour les enquêtes très longues nécessitant de poser de très nombreuses questions, souvent de différentes manières.

Avec le data mining, on cherche des résultats qui confirment les attentes spécifiques concernant un programme, mais on ignore les résultats qui ne confirment pas ces attentes. Par exemple, un agent de programme peut croire fermement qu'un programme de formation a un impact positif sur les jeunes. Une fois que l'agent reçoit les données de l'évaluation, celui-ci remarque qu'il y a une augmentation statistiquement significative du temps travaillé, mais le revenu moyen de ces jeunes n'a statistiquement pas évolué. Rapporter l'augmentation du temps travaillé en omettant le fait qu'il n'y a eu aucun changement des revenus relève du data mining.

Le data mining peut survenir de deux façons. Premièrement, lorsqu'on ignore des éléments probants qui vont à l'encontre de nos croyances, et qu'on ne rapporte que ceux qui les confirment. Deuxièmement, en cas d'anomalie statistique. En matière de statistiques, il y a toujours une chance pour qu'une variable soit jugée significative. En effet, dans au moins 5 % des cas, un élément sera jugé significatif alors qu'il ne l'est pas. Lorsqu'un évaluateur relève 100 éléments d'information, au moins cinq seront jugés importants, alors qu'ils ne le sont pas. Si le chercheur se concentre seulement sur ces cinq éléments, et ne reporte que ces cinq éléments là, les résultats seront biaisés.

Une évaluation d'impact peut ne révéler aucun élément statistiquement significatif. Mais en tenant compte de l'hétérogénéité possible, il y a fort à parier qu'en raison d'aléas statistiques, les chercheurs pourront trouver quelque impact sur un groupe. Pour éviter le data mining, on doit identifier l'ensemble des résultats d'intérêt avant de procéder à l'analyse, et ne pas omettre le moindre résultat, y compris ceux où aucun impact n'apparaît. De cette manière, on peut embrasser l'ensemble de la situation.

Interprétation des résultats

Qualité de la mise en œuvre : Les résultats dépendent beaucoup de la manière dont une intervention a été mise en œuvre. Le rapport final d'évaluation devrait donc examiner la qualité de cette mise en œuvre par le détail. Avoir une bonne connaissance de la manière dont a été mis en œuvre le programme est important, particulièrement lorsque les résultats de l'évaluation montrent un impact limité, voire négatif, dans la mesure où cela permet de distinguer les problèmes liés à la mise en œuvre de ceux liés à la conception de l'évaluation. Pour être en mesure d'interpréter correctement les

résultats d'une évaluation d'impact, il est nécessaire d'encadrer celle-ci strictement, du point de vue du suivi, du processus d'évaluation, et d'autres outils visant à en améliorer la qualité.

Généralisation des résultats : Théoriquement, l'évaluation d'impact a une validité externe, ce qui signifie qu'on peut généraliser les résultats à d'autres populations et d'autres situations. Si tel est le cas, cela dépendra beaucoup de la stratégie d'échantillonnage choisie pour l'évaluation. Plus l'échantillon est représentatif, plus on peut être confiant, et imaginer qu'un programme pourrait également fonctionner avec différents groupes, et à plus grande échelle. Ceci a des implications importantes en termes d'évolutivité et de reproduction de l'intervention. Il est prudent de garder à l'esprit que le passage du temps, les situations différentes, et les modes opératoires spécifiques d'un site à un autre pourront affecter de manière significative les effets du programme, en bien ou en mal. La prudence est donc toujours de mise lorsqu'il s'agit de tirer les leçons d'une évaluation, de les reproduire d'un programme à un autre, tout en ne perdant pas de vue qu'un suivi et une évaluation sont obligatoires pour continuer é apprendre et pour améliorer le programme.

La diffusion des résultats

Une fois que les résultats de l'évaluation d'impact sont connus, la dernière étape consiste à les diffuser auprès du personnel du programme ainsi qu'aux personnes intéressées, en dehors du programme.

Diffusion interne

La diffusion interne des résultats d'évaluation est la base de l'apprentissage organisationnel. Partager ces résultats avec le personnel du programme, et le reste de l'organisation, est l'accomplissement d'une des motivations essentielles qui touche à la conduite d'une évaluation : l'amélioration de la gestion du programme (voir note 1). Pour susciter l'intérêt et l'appropriation des objectifs du projet, il est préférable de commencer le processus de diffusion interne immédiatement après la fin de l'enquête de base, par exemple, par le partage et la présentation des conclusions de base. Le cas échéant, les résultats de l'évaluation doivent ensuite être diffusés auprès des cadres et dirigeants des pays concernés. Le rapport peut inclure une réflexion sur la manière dont les résultats peuvent influer la conception d'initiatives actuelles ou futures.

Diffusion externe

La diffusion doit aussi cibler les parties prenantes externes, comme les autorités locales, les ministères nationaux, les ONG locales et internationales, les universités (notamment les départements spécialisés dans le développement, l'économie, la santé publique), des organisations multilatérales (ONU, Banque mondiale, banques régionales de développement) ou les donneurs bilatéraux (par exemple, l'USAID, GIZ, DFID). En effet, les conclusions des évaluations d'impact sont généralement très demandées, en particulier celles relatives aux interventions visant à améliorer les moyens de subsistance des jeunes, car les informations rigoureuses sur ce qui fonctionne ou pas en la matière sont encore rares. Il y a de nombreuses façons de remettre des résultats d'évaluation à l'externe, les plans de diffusion prévoient généralement l'utilisation d'Internet ou le face à face (voir encadré 7.12.). Partager les résultats d'une évaluation peut avoir des répercussions dans le monde entier.

[Conseil]

Disposer de données sur la fréquentation liées au processus de suivi du programme peut s'avérer extrêmement utile. En effet, on est non seulement informé sur le nombre de jeunes inscrits au programme, mais aussi sur le fait de savoir quels services parmi ceux proposés ont été utilisés. Cela permet de distinguer entre les participants réguliers des irréguliers, et d'identifier les abandons en cours de programme (éventuellement remplacés). Si cette information n'est pas recueillie et analysée, il est probable que l'évaluation d'impact sous-estime l'efficacité du programme. Ces informations permettent aussi d'appréhender l'importance du *dosage* dans l'*évaluation*, par exemple, d'estimer la différence dans les résultats, entre quelqu'un qui a reçu 100 heures de formation et quelqu'un qui en a reçu seulement 50.

ENCADRÉ 7.12 Points de diffusion sélectionnés

Diffusion en ligne

- Site Internet des organisations
- Bulletins d'information
- Sites de partage connaissances (pour télécharger le rapport et les résultats)
  - Youth Employment Inventory http://www.youth-employment-inventory.org/
  - Youth Employment Network Groupsite http://yenclinic.groupsite.com
  - Eldis http://www.eldis.org/
  - Zunia http://zunia.org/
- Bases de données de recherche de documents
  - IZA Discussion Papers
  - http://www.iza.org/en/webcontent/publications/papers
  - Social Science Research Network
  - http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm
  - EconPapers
  - http://econpapers.repec.org/
- Les blogs et les réseaux sociaux

Diffusion en face à face

- Conférences thématiques
  - Global Youth Economic Opportunities Conference
  - http://www.youtheconomicopportunities.org
  - SEEP Annual Conference
  - http://www.seepnetwork.org/Pages/conference.aspx

Présentations

- Organisations internationales (Banque mondiale, BID, OCDE, OIT, UNICEF, PNUD, etc.)
- Donneurs bilatéraux (USAID, GIZ, DFID, AFD, etc.)
- Universités (locales ou internationales)

[Ressources en ligne]
Exemples de produits connexes

http://www.iyfnet.org/gpye-m&e-resource8

Produits connexes

Les notes d'orientation

Ces notes permettent de communiquer les résultats aux parties prenantes internes et externes. La note d'orientation (pas plus de quatre pages) contient les principales conclusions de l'évaluation en allant à l'essentiel. La note comprend également des graphiques et des tableaux et formule des recommandations.

Documents de travail

L'expert en évaluation à la possibilité de travailler avec l'équipe du programme pour rédiger des documents de travail et des articles destinés à être publiés dans des revues universitaires. Il pourra aussi présenter des résultats de recherche dans les universités et les instituts de recherche. Ces documents de travail peuvent alors être publiés et diffusés *via* les associations universitaires dont dépendent les enquêteurs. Le fait d'être cité dans des documents universitaires est un excellent moyen d'accroître la visibilité du programme et de susciter l'intérêt parmi les donateurs.

Résolution des problèmes

À l'instar de tous les types de programmes ou d'évaluations, on rencontre fréquemment des problèmes lors de mise en œuvre d'une évaluation d'impact. La liste suivante fournit des exemples de problèmes communs par étape, et propose des solutions pour les éviter ou les atténuer.

Préparation de l'évaluation

Mauvais programme à évaluer. Les évaluations d'impact dont les bénéfices et les contributions n'ont pas été clairement définis peuvent entraîner un gaspillage financier conséquent. Compte tenu des ressources limitées, il est important que les évaluations d'impact visent des interventions stratégiques et non encore testées avec un potentiel de répétition et d'évolution certains.

Objectifs irréalistes. De nombreuses interventions souffrent de ce que l'on pourrait appeler une « dérive de mission », par laquelle les objectifs du programme sont modifiés au fur et à mesure que le temps passe. Dans ces conditions, il est difficile d'établir des indicateurs d'évaluation valables. De même, si l'on définit des objectifs fantaisistes pour ce qui est des résultats attendus, l'évaluation risque de conclure qu'il n'y a pas d'impact. Il est important de rester réaliste au moment de définir les résultats recherchés et les objectifs d'apprentissage de l'évaluation.

Les influences externes. Même lorsque le plan d'évaluation a été défini et accepté, il se peut que des considérations politiques viennent entraver son déroulement. Ces facteurs externes peuvent précipiter ou retarder la mise en œuvre de l'évaluation, affectant la prestation de services et l'évaluation, par le retard ou l'incohérence du traitement, ou par effet de propagation sur le groupe de traitement et le groupe témoin, par exemple. Pour réduire l'influence des tiers, on peut convenir avec ces derniers d'un plan de mise en œuvre et d'évaluation (un protocole d'entente étant la meilleure solution), qu'il faudra réviser périodiquement.

Définir le calendrier et le budget de l'évaluation

Planning irréaliste. Lors de l'élaboration du calendrier et du budget, le risque principal est de sous-estimer le temps et les ressources nécessaires pour mener correctement à terme une évaluation d'impact. Il est fréquent qu'il y ait des retards dans la conception et la mise en œuvre du programme, ce qui entraîne à son tour une augmentation de la durée –et probablement des coûts– de l'évaluation. Par exemple, des consultants ou des membres de l'équipe ayant des postes clés peuvent ne plus être disponibles, entraînant ainsi des retards. La prudence en matière de budget et une gestion prévoyante du personnel sont essentielles.

Créer une équipe d'évaluation

Recrutement. Recruter une équipe d'évaluation d'impact efficace (ce qui implique diverses tâches, allant de la définition des mandats jusqu'à la localisation d'experts et d'instituts de sondage qualifiés) peut se révéler être un véritable défi. Sous-estimer l'expertise nécessaire dans les différentes étapes et recruter les mauvaises personnes peut entraîner d'importants retards et dépassements de coûts et, parfois, compromettre les résultats de l'évaluation. Il convient de s'assurer que les exigences pour chaque poste sont clairement définies à l'avance et que l'expert en évaluation et l'institut de sondage concernés les mènent bien à terme. Travailler avec des institutions comme les universités et les organismes de recherche, qui ont une expérience dans la conduite d'études de recherche de qualité, peut favoriser le soutien local facilitant ainsi une large acceptation des résultats finaux.

Changement de personnel. Les sociétés et instituts qui décrochent des contacts d'évaluation remplacent parfois le personnel aguerri par du personnel moins expérimenté. Cela peut être

évité grâce à des clauses contractuelles claires, des sanctions ou des mesures correctives.

Gestion de l'équipe d'enquête. Gérer une équipe d'enquête interne devient très vite compliqué. Lorsque la collecte de données doit être réalisée par le personnel du programme, il faut en appréhender les besoins dans leur intégralité, et veiller à ce qu'il y ait un encadrement et un suivi adéquats de l'équipe.

Élaborer un plan d'évaluation

Limites des données existantes. Lorsqu'on travaille avec des données secondaires, il est important de s'assurer qu'elles sont disponibles et de qualité. Les questionnaires déjà existants peuvent ne pas être très pertinents pour la nouvelle évaluation, ne pas cibler la population d'intérêt, ou avoir une taille d'échantillon trop petite pour pouvoir alimenter la nouvelle étude de manière adéquate. Avant de s'engager à n'utiliser que des données existantes, il est important d'en cerner les limites**.**

Programme et évaluation non connectés. Le manque de communication et de coordination entre l'organisation chargée de la mise en œuvre et l'investigateur principal peut pousser à choisir un concept d'évaluation non réalisable dans la pratique. Il faut faire en sorte que le personnel clé du programme soit associé à la planification de l'évaluation pour qu'elle soit adaptée au contexte opérationnel. Si ce genre de dysfonctionnement survient et qu'il est traité à temps, la meilleure solution consiste à trouver une méthode d'évaluation plus réaliste.

Biais de sélection. Identifier soigneusement l'échantillon et randomiser les participants à l'étude est la manière la plus simple et la plus sûre d'éliminer les biais de sélection. Si un biais de sélection est présent dans les données, une solution, bien qu'imparfaite, consiste à comparer les résultats chez le groupe traité à un échantillon apparié extrait d'un ensemble de données différent.

Élaborer et conduire un outil d'enquête

Mesuremania. Cibler de trop nombreux objectifs, et donc élaborer un questionnaire trop long, accroît souvent le coût de l'évaluation au-delà du budget de départ. Trop de questions rend plus l'enquête plus lourde, et peut entraîner la réduction du nombre et de la qualité des réponses. Supprimer les questions liées à des objectifs secondaires est un bon moyen d'atténuer le problème.

Essais insuffisants. La conduite des outils d'évaluation est une étape trop souvent ignorée. Cette étape est cruciale dans le processus d'évaluation, et ne saurait être éliminée, en particulier lorsque des jeunes sont interviewés, ce qui impose des défis supplémentaires dont la portée n'est pas perçue immédiatement. Si l'outil d'enquête n'est pas validé, les résultats pourraient être faussés, incomplets, voire trompeurs. Pendant la phase d'essai d'une enquête sur le terrain, il faut prendre le temps nécessaire pour s'assurer que les informations recueillies sont de premier ordre.

Ignorer l'éthique. Mener une enquête non approuvée par un IRB ou un comité d'éthique local peut conduire à un rejet massif des parties prenantes, et peut discréditer toute l'évaluation. Aussi, il convient de fournir au minimum une formation de base sur l'éthique à toutes les personnes impliquées par la collecte de données.

Mener une enquête et des analyses de base

Trouver des répondants. Il peut être difficile de localiser des jeunes pour les besoins de l'enquête. Dans ce cas, il est conseillé d'associer le personnel du programme local et les autres parties prenantes pour trouver ces répondants.

Qualité des données. Les instituts de sondage professionnels ne s'y entendent pas toujours très bien en matière d'évaluation d'impact, et par conséquent ne sont pas aussi qualifiés et fiables que l'on pourrait s'y attendre. Les intervieweurs peuvent falsifier ou mal enregistrer les

informations. De mauvaises méthodes de collecte de données ne sauraient être tolérées. Si des données de mauvaise qualité, voire falsifiées sont découvertes, il est important de le communiquer à l'institut de sondage, de lui signifier que cette situation est inacceptable, et que la collecte de données doit être recommencée en vue de garantir des normes élevées de qualité. Pour repérer ces dysfonctionnements et en réduire le nombre, il faut s'allouer les services d'un cabinet d'audit indépendant pour superviser la collecte de données. Le taux de vérification habituel d'un audit s'échelonne de 10 à 15 % des enquêtes, il permet de s'assurer que les répondants existent et que les données ont été collectées avec précision. Lorsque des incidents sont révélés, certains enquêteurs peuvent être affectés à d'autres tâches, voire licenciés.

Perte de données. Cela arrive quand les questionnaires remplis sont perdus, les ordinateurs ont été volés ou endommagés. Les données informatiques doivent toujours être sauvegardées. Sur le terrain, les questionnaires devraient être recueillis le plus tôt possible auprès des enquêteurs, si possible deux à trois fois par semaine, pour les protéger contre les pertes. Si des données sont entièrement perdues, il est préférable de recommencer la collecte de données. Cela signifie revoir les personnes déjà interrogées et leur expliquer qu'elles vont devoir l'être à nouveau. Recommencer la collecte de données peut s'avérer très fastidieux pour les répondants, sans oublier l'augmentation du coût du programme.

Saisie de données. La saisie de données doit être faite rapidement, dès que les questionnaires sont recueillis. Cela permet d'identifier les problèmes et de les corriger rapidement. En outre, les erreurs apparaissent souvent lors de la saisie des données. La plupart des logiciels de gestion de *données* incluent l'option de la double saisie (qui n'est pas établie d'office), avec laquelle chaque valeur est entrée deux fois. Les erreurs de transcription diminuent aussi grâce à l'utilisation de téléphones portables, de PDA, d'ordinateurs portables ou de tablettes numériques, pour la saisie des données.

Hypothèses erronées. Les principales hypothèses pour le modèle d'évaluation choisi pourraient ne pas être avérées. En utilisant de manière systématique les tests de vérification et de falsification (annexe 3), on peut détecter ces cas lors de l'analyse de base, et prendre des mesures précises, en modifiant notamment la stratégie d'évaluation. Pour réduire les possibilités que la conception choisie ne soit pas validée, il est important que le personnel du programme et de l'évaluation communiquent et restent en étroite collaboration, en s'assurant que l'enregistrement du programme et la collecte de données sont en phase avec les exigences d'évaluation.

Mener une enquête de suivi et des analyses

Attrition. L'attrition représente un problème important pour les évaluations et peut gravement altérer la pertinence des résultats. De toute évidence, il vaut mieux recourir à la prévention plutôt qu'à des mesures d'atténuation. Obtenir des informations de contact pertinentes lors de l'enquête de base, prévoir des mesures pour inciter les jeunes à participer à l'enquête, et utiliser les enquêtes de suivi peut aider à minimiser le phénomène d'attrition. Si le programme subit une forte attrition, malgré tous les efforts de prévention, il faudra recourir à la technique d'atténuation, qui consiste à sélectionner un échantillon aléatoire de personnes n'ayant pas été localisées pour mener des recherches très intensives à leur endroit. Ces unités, une fois localisées, représenteront de manière adéquate celles qui ne l'ont pas été. Enfin, dans la mesure où une certaine attrition est inévitable, il est possible d'en tenir compte lors de la définition de l'échantillon d'évaluation. Le fait de prévoir un échantillon de 10 à 20 % plus grand que nécessaire permet de disposer d'un nombre suffisamment large de réponses, et ainsi d'obtenir des résultats statistiquement significatifs même s'il y a déperdition (le biais potentiel résultant de l'attrition n'étant cependant pas compensé).

Non-conformité. En plus de l'attrition, il existe d'autres cas où des personnes ne sont pas en totale conformité avec les critères de sélection d'un programme. Il arrive, par exemple, que des jeunes sélectionnés pour suivre un programme de formation ne s'y rendent pas, tandis que d'autres, qui ont été affectés au groupe de comparaison le suivront. Une comparaison rigoureuse des résultats entre le groupe de traitement officiel et le groupe témoin faussera l'impact réel du programme. Tant que ces cas sont limités, et qu'on sait exactement qui dans le groupe de traitement et dans le groupe témoin a reçu une formation (grâce aux registres du programme), il est possible de corriger la non-conformité *via* des techniques statistiques, comme l'estimateur TOT (effet du traitement-sur-le-traité), que l'évaluateur sera à même de calculer.

Évaluation de la « boîte noire ». Le manque de connaissance sur la façon dont le programme à été mis en œuvre est un autre des problèmes fréquents du suivi d'impact. Cela conduit à des évaluations qui ne peuvent attribuer les changements observés (ou leur absence) à la conception ou la mise en œuvre du programme. Une solution possible consiste à intégrer les résultats du système de suivi et de compléter l'évaluation d'impact avec une évaluation de processus (voir aussi Méthodes mixtes à la note 8).

Diffusion des résultats

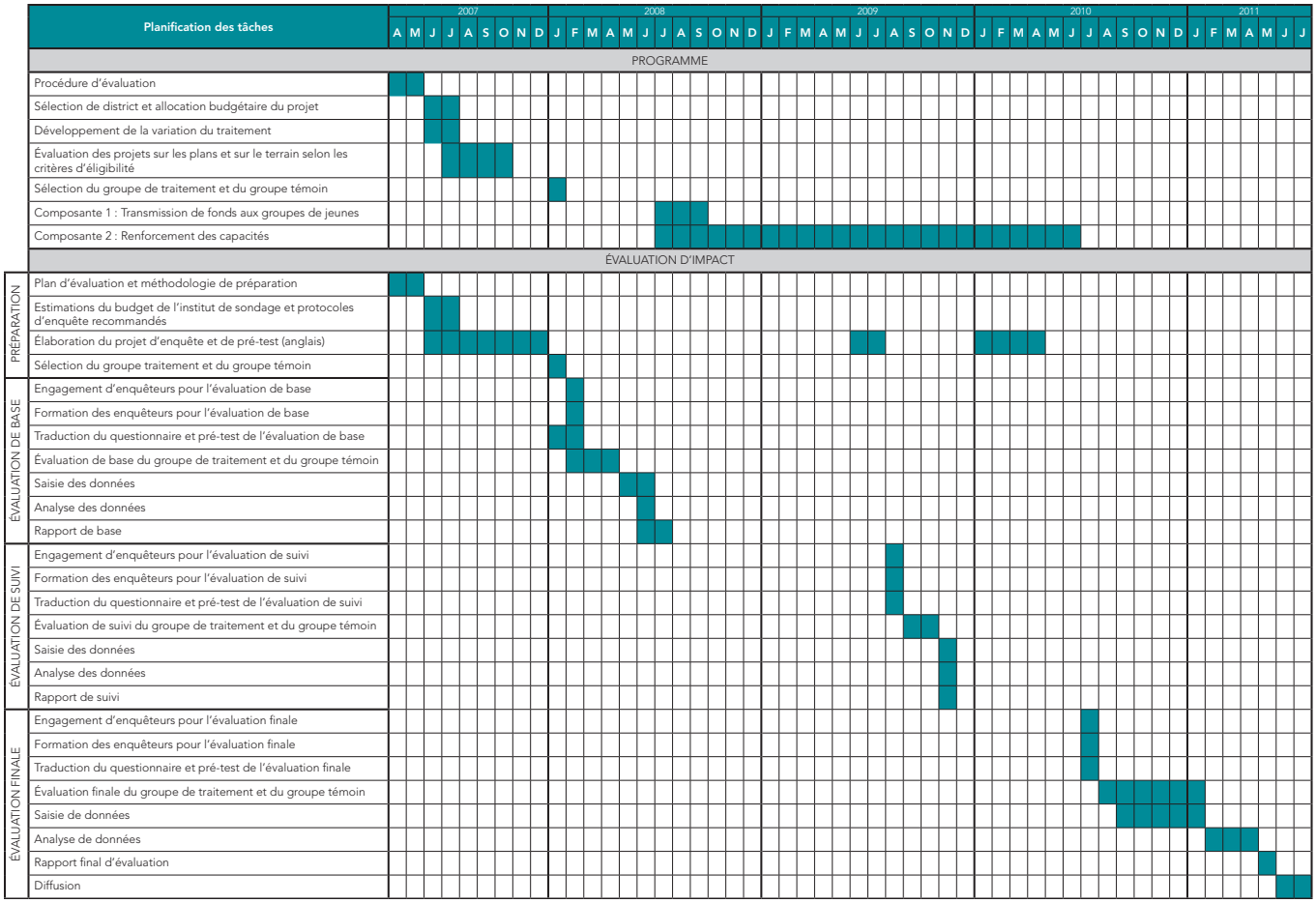
Utilisation limitée des résultats de l'évaluation. Si les résultats de l'évaluation ne sont pas suffisamment diffusés entre les parties prenantes internes et externes, alors les principaux objectifs de l'évaluation du programme, et plus largement dans le domaine des moyens de subsistance des jeunes, sont compromis. Pour surmonter ce problème, il faut définir une stratégie de diffusion dès le début de l'évaluation, et insister pour qu'au moins un membre du personnel du programme travaille en étroite collaboration avec l'équipe d'évaluation. Ainsi, au moins une personne clé du programme comprendra l'évaluation et sera bien placée pour mettre en œuvre certaines des conclusions du rapport.

Points essentiels

1. Mener une évaluation d'impact peut être coûteux et nécessiter beaucoup de temps, sans compter les nombreux problèmes qui peuvent se présenter. Il est donc essentiel de réunir une équipe de hautement qualifié, qui peut travailler à l'évaluation sur une période de temps prolongée.
2. Le plan d'évaluation est une étape importante de l'évaluation d'impact. Il présente la stratégie adoptée pour évaluer l'intervention, notamment la méthodologie de recherche, la taille de l'échantillon, le plan de collecte des données, ainsi que d'autres éléments.
3. Interviewer des enfants et des jeunes pose quelques problèmes particuliers, allant de l'obtention du consentement parental à l'utilisation d'un langage approprié. Aussi est-il préférable d'engager un expert en sondages. Les évaluations peuvent soulever des questions d'ordre éthique, il faudra donc soumettre le concept d'évaluation et l'enquête à un comité d'examen institutionnel (IRB).
4. Mener une enquête de base est fortement recommandé, car elle fournit des informations précieuses sur la conception des programmes et nous permet de tester la faisabilité du modèle d'évaluation choisi.
5. Le calendrier de collecte des données de suivi doit être bien conçu pour saisir des résultats d'intérêt, dont certains se matérialiseront sur le court terme, tandis que d'autres nécessiteront plusieurs années pour le faire.
6. Il est primordial de diffuser largement les résultats d'évaluations, qu'ils soient positifs ou négatifs. Le partage des résultats avec les parties prenantes internes, locales et internationales permet d'apprendre encore davantage et de recevoir des retours concernant l'intervention.

Étude de cas du NUSAF : Mise en œuvre de l'évaluation d'impact

L'évaluation du « Youth Opportunities Program », programme sous l'égide du NUSAF visant à améliorer les perspectives d'emploi des jeunes, a commencé en Juin 2007 et s'est achevée en mai 2011 avec la réalisation du rapport final d'évaluation. Le programme a distribué des fonds aux participants d'août à septembre 2008. L'évaluation comprenait une enquête de base qui a eu lieu début de l'année 2008, une enquête de suivi, menée fin 2009, et une enquête finale réalisée entre fin 2010 et début 2011. Chacune des enquêtes couvrait l'ensemble des participants.


Planification des tâches
2007
2008
2009
2010
2011
A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J
PROGRAMME
Procédure d'évaluation
Sélection de district et allocation budgétaire du projet
Développement de la variation du traitement
Évaluation des projets sur les plans et sur le terrain selon les critères d'éligibilité
Sélection du groupe de traitement et du groupe témoin
Composante 1 : Transmission de fonds aux groupes de jeunes
Composante 2 : Renforcement des capacités
ÉVALUATION D'IMPACT
PRÉPARATION
Plan d'évaluation et méthodologie de préparation
Estimations du budget de l'institut de sondage et protocoles d'enquête recommandés
Élaboration du projet d'enquête et de pré-test (anglais)
Sélection du groupe traitement et du groupe témoin
ÉVALUATION DE BASE
Engagement d'enquêteurs pour l'évaluation de base
Formation des enquêteurs pour l'évaluation de base
Traduction du questionnaire et pré-test de l'évaluation de base
Évaluation de base du groupe de traitement et du groupe témoin
Saisie des données
Analyse des données
Rapport de base
ÉVALUATION DE SUIVI
Engagement d'enquêteurs pour l'évaluation de suivi
Formation des enquêteurs pour l'évaluation de suivi
Traduction du questionnaire et pré-test de l'évaluation de suivi
Évaluation de suivi du groupe de traitement et du groupe témoin
Saisie des données
Analyse des données
Rapport de suivi
ÉVALUATION FINALE
Engagement d'enquêteurs pour l'évaluation finale
Formation des enquêteurs pour l'évaluation finale
Traduction du questionnaire et pré-test de l'évaluation finale
Évaluation finale du groupe de traitement et du groupe témoin
Saisie de données
Analyse de données
Rapport final d'évaluation
Diffusion

Source : Blattman, Fiala, and Martinez (2011).

À lire

Baker, J. 2000. *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*, Washington, DC: The World Bank. (Chapitre 2 en rapport avec cette note)
http://siteresources.worldbank.org/INTISPMA/Resources/handbook.pdf

Bamberger, M., Rugh, J., and Mabry, L. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks: Sage Publications. (Voir chapitre 3–8.)
http://realworldevaluation.org/

Gertler, P., Martinez, S., Premand, P., Rawlings, L., and Vermeersch, C. 2011. *Impact Evaluation in Practice*. Washington, DC: The World Bank. (Voir chapitres 10–13.)
http://www.worldbank.org/ieinpractice

NOTE 8 : Augmenter la pertinence de l'évaluation d'impact

Une évaluation d'impact quantitative type est tout naturellement d'une grande valeur pour notre programme et pour l'organisation, mais il existe diverses options visant à améliorer encore la qualité de l'analyse et augmenter la pertinence des résultats. D'une part, les évaluations d'impact peuvent répondre à diverses questions de recherche, y compris, mais bien au-delà, à celles relatives aux effets moyens du programme. D'autre part, lorsqu'elles sont combinées avec des outils qualitatifs, les évaluations d'impact peuvent être beaucoup plus enrichissantes que lorsqu'elles s'appuient uniquement sur des méthodes quantitatives. Enfin, les résultats d'une évaluation d'impact peuvent être optimisés afin d'obtenir des analyses plus poussées, par exemple pour estimer les avantages globaux du programme avec ses coûts totaux (en menant une analyse coûts-bénéfices). La note finale présente succinctement ces outils, pour que les praticiens puissent tirer le meilleur parti de leur évaluation de l'impact.

Mesure de divers impacts

L'évaluation des impacts de base répond à la question « Le programme a-t-il fonctionné ? », en d'autres termes a-t-il atteint les résultats d'intérêt tels que définis dans le programme et les objectifs d'apprentissage ? Il est important de savoir si le programme a eu un impact, dans son ensemble, mais il convient aussi de poser d'autres questions.

Il est sans doute utile d'avoir une image plus nuancée de l'impact réel du programme. Cela sous-entend une meilleure appréhension des questions suivantes :

- Les résultats varient-ils selon les différents groupes de bénéficiaires (par exemple, les garçons en profitent, mais pas les filles) ?
- Quels sont les impacts de l'intervention à court et à long terme ?
- Le programme a-t-il des retombées positives ou négatives ? Y a t-il des résultats prévus ou imprévus au-delà du groupe cible existant ?

Deuxièmement, il peut être également intéressant d'avoir recours à des modèles transversaux (*crosscutting designs,* CCD en anglais) pour observer la variation de l'efficacité du programme en fonction du modèle choisi. Le recours aux modèles transversaux permet de répondre aux questions suivantes :

- Y a-t-il des concepts de programmes plus efficaces que d'autres ? Pour le savoir, nous pouvons comparer les types d'interventions (accorder des subventions de démarrage à de jeunes entrepreneurs, plutôt qu'un prêt à la création d'entreprise, par exemple), ou tester la combinaison la plus efficace des composantes du programme (une formation simple, une formation avec stage et une formation avec stage et mentorat)
- Quel est le « dosage » le plus efficace des activités proposées par le programme ? Par exemple, doit-on offrir 20, 50, ou 100 heures de formation ?

Si elles sont bien conçues, les évaluations d'impact peuvent répondre à ces questions, même s'il est difficile d'avoir toutes les réponses en menant une seule évaluation d'impact. Comme chaque intervention a des priorités et des objectifs d'apprentissage différents, on peut concevoir l'évaluation d'impact de telle sorte qu'elle apporte des réponses aux questions les plus pertinentes concernant le programme. Un éventail plus large de questions permettra d'améliorer la pertinence des résultats de l'évaluation. Cependant, il est tout aussi important de comprendre que des données supplémentaires sont nécessaires pour évaluer ces questions et les modèles transversaux (voir tableau 8.1.).

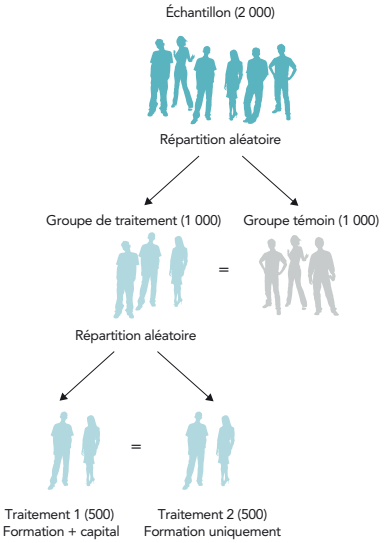
TABLEAU 8.1 Types de questions concernant les évaluations d'impact

Question	Description	Données supplémentaires requises	Résultat de l'évaluation de l'échantillon et interprétation
Quel est l'impact global du programme sur les résultats A, B et C dans le groupe X, dans un contexte Y ?	Ce sont les questions type d'une évaluation d'impact.	• s. o. (collecte de données type basée sur la méthode choisie)	L'impact moyen du programme de formation sur le revenu des jeunes est de + 20 USD par mois. Le programme a un impact positif sur le revenu.
Les résultats varient selon les groupes de population ?	Les interventions touchent souvent les groupes différemment (hétérogénéité des impacts). Mesurer uniquement l'impact moyen peut masquer ces différences, il faut donc fractionner les impacts par groupe de population.	• Informations sociodémographiques des participants et du groupe témoin (âge, sexe, niveau de revenu, etc.) • Pour être en mesure de ventiler les résultats, le nombre de personnes couvertes par l'évaluation (la taille de l'échantillon) doit augmenter avec chaque catégorie d'informations qui doit être analysée	L'augmentation moyenne des revenus est de 40 s pour les garçons, et 0 USD pour les filles. Les jeunes plus âgés sont plus avantagés que les plus jeunes (30 USD contre 10 USD en moyenne). Par conséquent, l'efficacité du programme n'est pas la même pour tous les participants. Il faut comprendre pourquoi les groupes bénéficient différemment des interventions, et adapter, dans la mesure du possible, les orientations stratégiques et la conception du programme pour répondre au besoin de certaines populations.
Quels sont les impacts de l'intervention à court et à long terme ?	Les résultats ne son pas toujours constants dans le temps. Certains effets à court terme peuvent s'amenuiser, tandis que d'autres à long terme n'apparaîtront qu'au bout de plusieurs années, une fois l'intervention terminée.	• Données étalées sur une longue période (en pratique, cela signifie qu'il faut suivre le groupe de traitement et le groupe témoin pendant plusieurs années)	À la fin du programme, on observe un revenu mensuel moyen pour le groupe traité de – 5 USD (une perte donc) par rapport au groupe témoin. Deux ans après la fin du programme, l'augmentation moyenne du revenu mensuel pour le groupe de traitement a été de 20 USD. Les participants à la formation n'ont pu travailler autant que leurs pairs pendant cette période, aussi ont-ils subi une baisse de leurs revenus. Cependant, à long terme, la formation a fini par porter ses fruits, et les participants ont pu obtenir des revenus supérieurs à celui des personnes non formées. Rechercher seulement des résultats à court terme peut être trompeur.
Le programme a-t-il eu des retombées?	Le programme peut avoir des retombées indirectes (positives ou négatives) sur les non participants.	• Données au-delà du groupe de traitement et du groupe témoin (membres de famille et de leur communauté)	Non seulement les participants ont un revenu moyen plus élevé de 20 USD, mais leurs voisins ont également connu une augmentation de leur revenu mensuel de 5 USD. Les participants ont apparemment transmis les connaissances aux autres.
La conception du programme A, est-elle plus efficace que celle du B, et inversement ?	Les questions relatives à la conception du meilleur programme possible sont souvent ambigües. En effet, doit-on avoir recourt à d'autres types d'intervention ou doit-on combiner les composantes du programme différemment ?	• Plusieurs groupes de traitement (un groupe régi par la conception de type A, un autre groupe par la B, etc.) • Le nombre de personnes couvertes par l'évaluation doit être suffisamment important pour permettre de créer plus d'un groupe de traitement et plus d'un groupe témoin	L'augmentation moyenne des revenus est de 5 USD pour ceux qui ont reçu une formation, et de 30 USD pour ceux qui ont reçu une formation et un stage. Ainsi, fournir une expérience de travail pratique en plus de la formation semble améliorer considérablement l'impact de l'intervention.
Quel est le « dosage » le plus efficace des activités proposées par le programme ?	Jouer la surenchère n'est pas toujours la panacée. Il s'agit de trouver l'équilibre entre le volume des prestations offertes, pour optimiser l'impact de l'intervention, et le coût de celles-ci.	• Plusieurs groupes de traitement (un groupe régi par la conception de type A ; un autre groupe, par la B, etc.) • Le nombre de personnes couvertes par l'évaluation doit être suffisamment important pour permettre de créer plus d'un groupe de traitement et plus d'un groupe témoin.	L'augmentation moyenne du revenu est de 0 USD pour ceux qui ont reçu un mois de formation, 20 USD pour trois mois, et 20 USD pour six mois. On remarque qu'un seul mois de formation n'est pas suffisant, six mois de formation n'ont amené aucun avantage supplémentaire par rapport à trois mois de formation. La durée optimale de la formation semble être d'environ trois mois.

Les modèles transversaux permettent non seulement d'évaluer l'impact global du programme, mais aussi ses caractéristiques spécifiques, et pourquoi celles-ci fonctionnent ou pas. Un programme peut fournir, par exemple, une formation professionnelle et entrepreneuriale, dans le domaine de la menuiserie ou la couture, agrémentée d'une petite aide financière à la création d'entreprise. L'octroi de subventions en espèces pourrait être coûteux ou politiquement difficile. Aussi, le directeur du programme peut-il s'interroger sur la nécessité d'octroyer un capital de démarrage, ou sur le fait de savoir si les participants sont capables de mettre en œuvre leur formation sans ce capital. Dans ces exemples, les modèles transversaux permettent de déterminer la meilleure conception de programme.

Les modèles transversaux nécessitent au moins deux groupes de traitement recevant différentes combinaisons ou « dosages » du programme. Ces deux groupes peuvent ensuite être comparés lors de l'évaluation finale. La différence entre les deux groupes correspond à l'impact spécifique de chaque combinaison. En reprenant l'exemple précédemment évoqué, l'évaluation peut avoir un échantillon de 2 000 participants répartis par répartition aléatoire entre le groupe de traitement et le groupe témoin. Le groupe de traitement peut ensuite être randomisé en deux sous-groupes de traitement (traitement 1 et 2). Dans le traitement 1, les participants bénéficient de la formation et du capital de démarrage. Dans le traitement 2, les participants bénéficient seulement de la formation (voir figure 8.1).

FIGURE 8.1 Schéma d'une évaluation d'impact avec recours aux modèles transversaux



La détermination de l'impact de l'octroi d'un capital de démarrage résulte de la comparaison entre les sous-groupes de traitement 1 et 2, lors de l'évaluation finale. L'impact de la formation est déterminé en comparant le groupe de traitement de 2 à ceux du groupe de comparaison.

Utilisation de méthodes mixtes d'approches

Il est important de ne pas perdre de vue les limites des méthodes quantitatives d'évaluation d'impact. Si elles sont utilisées de manière isolée, on risque de ne pas saisir la complexité des résultats du programme, et d'interpréter de manière inadéquate les impacts identifiables. Afin de comprendre correctement la dynamique d'une intervention, et de pouvoir expliquer les raisons pour lesquelles celle-ci fonctionne, il est important qu'un suivi et une évaluation du processus encadrent strictement

les techniques d'évaluation d'impact. Globalement, nous pensons que l'utilisation de méthodes mixtes, c'est-à-dire l'adoption explicite de méthodes quantitatives et qualitatives lors de la conception de l'évaluation d'impact, peut améliorer considérablement la connaissance de nos programmes.

Comme le soulignent Bamberger, Rao, and Woolcock (2010, pp. 6–7) et Leeuw and Vaessen (2009), l'utilisation de méthodes mixtes peut renforcer l'évaluation d'impact quantitative de plusieurs manières :

- Généralement, les évaluations d'impact quantitatives ne relèvent pas d'informations sur la qualité de la mise en œuvre du programme. L'appréciation du processus de mise en œuvre du programme est primordiale pour comprendre comment cette mise en œuvre peut affecter les résultats de l'intervention. En outre, cette appréciation permet d'interpréter convenablement les conclusions de l'évaluation et de savoir si des résultats décevants sont dus à des faiblesses dans la conception du programme ou à la mise en œuvre de celui-ci. Un bon suivi est déterminant pour obtenir une évaluation efficace ; il peut être complété par d'autres outils d'analyse des processus, comme les interviews avec des informateurs clés, les observations directes de participants, et les groupes de discussion.
- L'utilisation de méthodes qualitatives peut aider à comprendre comment et pourquoi les effets de l'intervention ont pu varier en fonction des populations cibles. Même si les techniques quantitatives sont conçues pour évaluer l'hétérogénéité d'impact dans les groupes concernés, celles-ci ne peuvent fournir d'explications précises nous permettant de comprendre le pourquoi cette hétérogénéité.
- Les conceptions quantitatives ne sont en principe pas capables de saisir la variété de situations locales qu'implique la mise en œuvre de chaque programme. Les méthodes mixtes peuvent fournir une analyse détaillée du contexte et enregistre les différences dans la qualité ou la vitesse de mise en œuvre à travers les sites du programme. Ces informations qualitatives peuvent expliquer les différences potentielles dans les résultats de programmes en fonction des zones géographiques.
- Les résultats d'interventions relatives au développement des moyens de subsistance des jeunes (santé mentale, autonomisation, ou les relations au sein des ménages) sont souvent complexes et multidimensionnels, et ne peuvent être appréhendés avec des méthodes quantitatives. Les méthodes mixtes permettent le suivi des indicateurs qualitatifs, et fournissent une analyse d'études de cas choisies visant à mieux comprendre la dynamique et les résultats de l'intervention. Par exemple, les entretiens qualitatifs structurées et semi-structurés au cours desquels les participants sont libres d'évoquer de vraies histoires, qui ne relèvent pas de la catégorie des informations quantifiables, peuvent aider à comprendre les effets d'un programme. Les méthodes qualitatives sont également mieux adaptées pour recueillir des informations sur des sujets sensibles, comme la santé sexuelle et reproductive ou la violence.
- Les méthodes qualitatives permettent avant tout d'identifier des indicateurs appropriés. Un groupe de discussion peut donner, par exemple, des informations importantes sur les préoccupations des bénéficiaires, et ce qu'ils attendent-ils de l'intervention.

Incorporer des éléments qualitatifs dans une évaluation d'impact peut prendre plusieurs formes, notamment les enquêtes à questions ouvertes, les sélections d'entretiens approfondis et études de cas, les discussions de groupe, les outils participatifs, comme le « Most Significant Change », en d'autres termes la technique du « changement le plus significatif » (voir Davies et Dart 2005), la méthode de l'observateur participant, etc. Pour en apprendre davantage sur les approches participatives du suivi et évaluation, consulter Catley *et al.* (2010), Sabo Flores (2008), Powers et Tiffany (2006), et Gawler (2005).

Les données qualitatives seules ne sont pas adaptées pour identifier les impacts du programme. L'utilisation de méthodes mixtes permet, par conséquent, de compenser les faiblesses des outils d'évaluation qualitatifs et quantitatifs. Aussi, la combinaison des deux types d'outils permet-elle de concevoir une évaluation plus aboutie (encadré 8.1. fournit un exemple). L'utilisation combinée de plusieurs méthodes de recherche augmente la crédibilité et la validité de nos résultats.

Toutefois, il est important de noter que l'utilisation de méthodes mixtes peut entraîner une augmentation de la durée et des coûts, ainsi que des défis logistiques supplémentaires. En outre, il est fréquent que la combinaison entre les différentes disciplines professionnelles associées aux chercheurs rende long et difficile le recrutement d'une équipe multidisciplinaire. (Bamberger, Rao, Woolcock 2010, p. 17).

ENCADRÉ 8.1 Exemple d'évaluation à partir de méthodes mixtes

Dans une évaluation du programme d'études *Our Nation* de Junior Achievement (JA), Jeunes entreprises dans les pays francophones, les évaluateurs ont combiné des méthodes de recherche quantitatives et qualitatives. *Our Nation* est l'un des nombreux programmes destinés aux écoles élémentaires que JA diffuse à l'échelle de la planète. Ces programmes se composent d'une série de séances pédagogiques à l'attention d'élèves de 9 à 11 ans, qui abordent des problématiques liées à l'entrepreneuriat, aux ressources nécessaires pour réaliser un projet d'entreprise, et à la mondialisation. D'une part, l'évaluation s'est appuyée sur une conception expérimentale avec répartition aléatoire des élèves dans le groupe de traitement et le groupe témoin. Les unités du groupe témoin étaient originaires des mêmes états et régions que les élèves du groupe de traitement, mais un tirage au sort a été effectué pour déterminer qui bénéficierait du programme une fois l'évaluation accomplie. D'autre part, les évaluateurs ont mené plusieurs études de cas, sollicitant l'aide d'enseignants, de bénévoles, des enquêteurs de JA, des groupes de discussion d'étudiants, et en recourant à l'observation en classe.

Les résultats de l'évaluation quantitative ont démontré certains effets positifs sur la connaissance des élèves concernant l'entrepreneuriat et la mondialisation, mais aucun effet sur le niveau d'engagement scolaire et l'acquisition des compétences du XXIe siècle. En outre, les outils qualitatifs ont permis une compréhension en profondeur des mécanismes en œuvre. D'une part, les outils qualitatifs ont confirmé une mise en œuvre de bonne qualité, la majorité des séances ayant été exécutées conformément aux recommandations de JA. Les niveaux de satisfaction des étudiants, es enseignants et des bénévoles étaient élevés. Cependant, ces outils ont également souligné des carences pouvant compromettre le bon déroulement du programme, par exemple, le temps dévolu aux bénévoles n'était pas suffisant pour assurer l'intégralité de leurs missions. Enfin, les résultats de l'évaluation qualitative ont suggéré certains moyens visant à améliorer la mise en œuvre du programme, notamment augmenter la durée des sessions pédagogiques, un effort sur la réductions des difficultés liées vocabulaire utilisé, et fournir plus d'orientation pédagogique aux bénévoles.

Source : RMC Research (2009).

Analyse coûts-avantages et coût-efficacité

Dans de nombreux cas, les organisations peuvent utiliser différentes stratégies pour faire face au même problème. Par exemple, pour accroître les perspectives d'emploi on peut tenter d'améliorer les services d'orientation professionnelle ou la formation. Même si l'on ne choisit qu'une seule stratégie, il est possible de mettre en œuvre le programme selon différentes approches, notamment en ayant recours à des prestataires de formation publics ou privés. Si ces diverses stratégies de programmation et de mise en œuvre garantissaient le même impact, à savoir, si les deux permettaient d'accroître les perspectives d'emploi de 50 % trois mois après l'intervention, serions-nous également satisfaits si l'on choisissait une approche plutôt que l'autre ? Probablement pas, car il ne suffit pas de savoir qu'une intervention est viable, ni pour qui, ni dans quel contexte elle l'est ; il faut aussi en connaître les coûts.

En effet, une estimation réaliste des coûts permettra de répondre aux questions suivantes :

- Comment choisir une alternative ? Quel programme offre le meilleur rapport coût/efficacité pour un niveau d'impact donné ?
- Serions-nous capables d'étendre l'impact de l'intervention ? Si les coûts sont élevés, il est peu probable que nous soyons en mesure d'atteindre un grand nombre de bénéficiaires.
- Une intervention vaut-elle mieux que pas d'intervention du tout ? Si l'ensemble des avantages générés par le programme est en deçà de son coût total, peut-être vaudra-t-il mieux dépenser ces ressources autrement.

Outils analytiques

Les deux outils couramment utilisés pour répondre aux questions ci-dessus évoquées sont : l'analyse coût-efficacité (ACE) et l'analyse coûts-avantages (ACA) :

L'**analyse coût-efficacité** identifie le coût total d'un programme, en comparant des ratios de coût par unité de résultat (par exemple, 500 dollars E.-U par personne formée, par emploi créé, par personne sauvée de l'infection VIH / SIDA, etc.). L'ACE présente le résultat obtenu par dollar dépensé, en identifiant ainsi l'attribution de ressources la plus efficace en comparaison à d'autres types de programmes (voir tableau 8.2.).

TABLEAU 8.2 Estimation du coût-efficacité pour le programme *Jóvenes*

Pays	Programme	Coût par participant (en 2005)
Argentine	Proyecto Joven	1 159 USD
Chili	Chile Joven	825 –1 051 USD
Pérou	PROJoven	697 USD

Source : Betcherman *et al.* (2007).

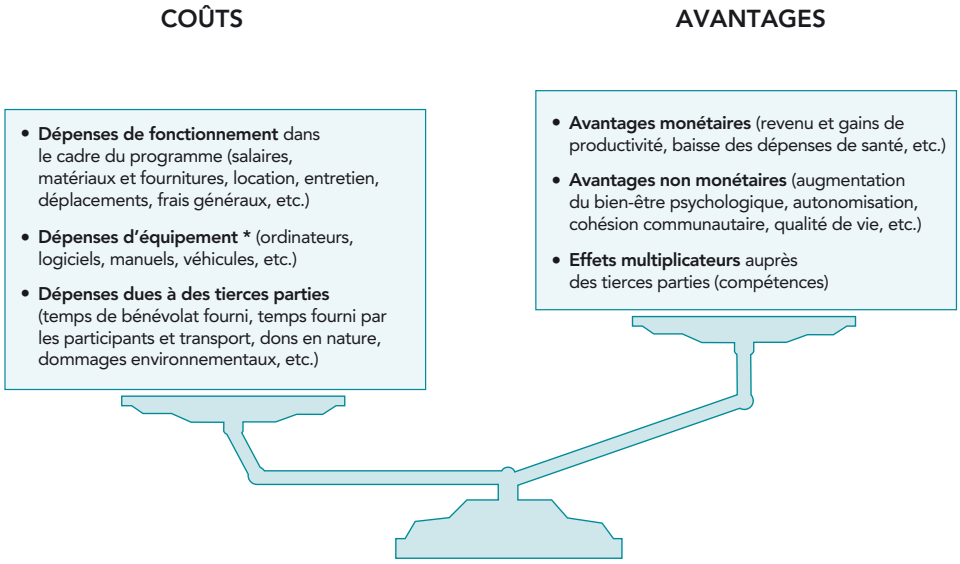
L'**analyse coûts-avantages** identifie et quantifie le coût total d'un programme, et met ce coût en relation avec le total des avantages du programme mesuré en dollars. Une fois connus les bénéfices et les coûts nets de l'intervention, il est possible de calculer le ratio coûts-avantages et de déterminer le retour social sur investissement. Par exemple, le ratio bénéfices-coûts est de 2:1 si les bénéfices nets sont de 1 000 dollars E.-U par personne, et les coûts nets sont de 500 dollars E.-U. Globalement, l'ACA vise à déterminer si les avantages compensent les coûts, c'est-à-dire si la société est gagnante ou perdante après avoir fait cet investissement.

L'ACE et l'ACA peuvent être utilisées avant, pendant ou après le programme. Toutefois, seule une analyse rétrospective fournira aux praticiens des informations complètes sur les coûts et les avantages réels, qui leur permettra de déterminer le succès de l'intervention. L'évaluation d'impact est une condition nécessaire pour disposer d'une estimation fiable des avantages directs et indirects du programme.

Saisir l'ensemble des avantages et des coûts

Pour mener des analyses coût-efficacité et coûts-avantages, il faut saisir, quantifier et comparer tous les coûts connus (et pour l'ACA les avantages connus) du programme pour les personnes directement ou indirectement concernées : l'organisation, les bénéficiaires du programme, le(s) gouvernement(s), etc. (voir figure 8.2).

FIGURE 8.2 Comparaison des coûts et bénéfices


COÛTS
AVANTAGES
• Dépenses de fonctionnement dans le cadre du programme (salaires, matériaux et fournitures, location, entretien, déplacements, frais généraux, etc.)
• Dépenses d'équipement * (ordinateurs, logiciels, manuels, véhicules, etc.)
• Dépenses dues à des tierces parties (temps de bénévolat fourni, temps fourni par les participants et transport, dons en nature, dommages environnementaux, etc.)
• Avantages monétaires (revenu et gains de productivité, baisse des dépenses de santé, etc.)
• Avantages non monétaires (augmentation du bien-être psychologique, autonomisation, cohésion communautaire, qualité de vie, etc.)
• Effets multiplicateurs auprès des tierces parties (compétences)

**Achat de biens dont l'usage dépasse une année.*

Le retour social sur investissement (SROI) est une variante de l'ACA, qui compare les avantages extra-financiers aux ressources investies. Il assigne des valeurs financières approximatives, les « proxys », aux résultats identifiés par les parties prenantes, qui n'ont généralement pas de valeur sur le marché. Pour en savoir plus, consulter le site Internet du Réseau SROI : http://www.thesroinetwork.org/

La tentative d'attribuer une valeur monétaire à de nombreux avantages immatériels peut se révéler difficile et subjective, voire symboliser un défi important, surtout pour l'ACA. Par conséquent, l'ACA est généralement considérée comme la plus utile lorsqu'il est question de quantifier en termes monétaires plusieurs types d'avantages et de consensus (J-PAL 2011).

Calcul des avantages nets pour les participants

Les avantages d'un programme peuvent être évalués d'après son impact sur chaque participant. Par exemple, imaginons un programme de formation professionnelle qui permette d'augmenter le revenu annuel par personne de 100 dollars E.-U en moyenne. Dans de nombreuses régions d'Afrique subsaharienne, il s'agit d'une importante somme d'argent, ce qui serait un grand succès pour le programme.

Supposons que la mise en œuvre du programme coûte un million de dollars, dont 400 000 dollars E.-U sont attribués à la formation et 600 000 dollars E.-U aux frais généraux, notamment les salaires du personnel. Si le programme est prévu pour 1 000 personnes, sa mise en œuvre coûte donc 1 000 dollars E.-U par personne, dont 400 dollars E.-U pour la formation et 600 dollars E.-U pour les frais généraux. Le jeu en vaut-il la chandelle ?

La réponse à cette question dépend de trois critères.

Premièrement, **l'impact du programme doit égaler ou dépasser celui consistant à donner un montant d'espèces à des particuliers équivalant au coût du programme.** En reprenant l'exemple ci-dessus, l'impact doit être comparé aux conséquences que produirait un don de 1 000 dollars E.-U en espèces à chaque personne. Deux scénarios sont alors possibles. Premièrement, une personne utilise les 1 000 dollars E.-U pour financer une formation. Le coût de la formation ne coûte toujours que 400 dollars E.-U par personne, et occasionne un rendement annuel de 100 dollars E.-U par personne. Cet individu se retrouve donc avec une somme de 600 dollars E.-U à utiliser comme bon lui semble, sa situation est donc plus avantageuse que dans le programme. Dans un second scénario, une personne peut utiliser l'argent pour autre chose que la formation, qui est moins

utile pour elle à long terme, contrairement aux cigarettes ou d'autres biens de consommation non essentiels. Dans ce dernier cas, le programme vaut la peine d'être mené.

Deuxièmement, **le programme doit avoir un rendement égal ou supérieur à l'exécution d'autres programmes.** Un autre programme aurait-il pu produire le même ou un plus grand impact par personne, tout en coûtant moins cher ? Cette question exige une comparaison des résultats entre les différentes options des programmes. Le programme qui a le plus d'impact, et qui coûte le moins cher est donc le programme qui doit continuer à fonctionner.

Enfin, **la valeur actuelle nette de rendement devrait être plus élevé que le coût du programme.** La valeur actuelle est une manière de penser à la valeur de l'argent aujourd'hui par rapport à sa valeur future. En utilisant le même exemple, nous prenons la valeur de l'argent obtenu annuellement (les 100 dollars E.-U de rendement par personne après la formation) et l'ajustons pour une période de temps déterminé en fonction du taux d'escompte, qui dans la plupart des cas équivaut au taux d'intérêt local. La valeur actuelle nette est simplement la somme de la valeur actuelle ajustée sur une période de temps. Le tableau 8.3. ci-après illustre cette démonstration, en utilisant un taux d'intérêt de 20 %, qui est un taux commun dans de nombreux pays en développement.

TABLEAU 8.3 Valeur actuelle et valeur actuelle nette pour un rendement annuel de 100 dollars E.-U

Année	Valeur actuelle
0	100 USD
1	83 USD
2	69 USD
3	58 USD
4	48 USD
5	40 USD
6	33 USD
7	28 USD
8	23 USD
9	19 USD
10	16 USD
Valeur actuelle nette	517 USD

Dans cet exemple, 100 dollars E.-U d'aujourd'hui sont évalués 100 dollars E.-U. Cependant, au taux d'intérêt actuel, ces 100 dollars E.-U de revenu supplémentaire (rendement) d'aujourd'hui vaudront seulement 16 dollars E.-U dans dix ans. La valeur actuelle nette de ces 100 dollars E.-U représente 517 dollars E.-U, sur une période de 10 ans. Sur l'ensemble de la durée de vie d'un participant, la valeur actuelle nette s'élèvera à plus de 600 dollars E.-U. Dans ces conditions, un rendement annuel de 100 dollars E.-U par personne équivaut à un rendement maximal de 600 dollars E.-U par personne pendant toute la vie.

Selon ces critères, nos dépenses en dollars d'aujourd'hui pour le programme (1 000 dollars E.-U) sont supérieures aux bénéfices qu'en retirent les individus, même si le coût de la formation (400 dollars E.-U) est inférieur à la valeur actuelle nette de la formation (517 dollars E.-U). Par conséquent, à moins que les intéressés ne soient priés de prendre la formation à leur charge, le budget alloué aux frais généraux n'étant alors plus de mise, ou à moins de réduire considérablement ces frais généraux, l'utilité de la formation ne justifie pas à elle seule l'existence du programme.

Calcul des avantages nets pour la société

Afin d'évaluer les avantages nets pour la société, nous devons tenir compte des effets d'entraînement. Ces effets se réfèrent à des impacts positifs ou négatifs du programme sur les individus n'y étant pas directement impliqués. Deux types d'effets d'entraînement vont nous intéresser : les effets multiplicateurs et les effets prix et quantité.

Effets multiplicateurs

Les effets multiplicateurs se manifestent lorsque les participants à un programme transmettent leurs compétences à des tiers n'y étant pas officiellement associés. Un individu ayant reçu une formation en menuiserie peut former son beau-fils, par exemple. L'évaluation d'impact peut seulement mesurer les effets du programme sur le participant, mais pas ceux sur le beau-fils.

En reprenant l'exemple ci-dessus, l'impact sur le menuisier se matérialise par une augmentation de son revenu de 100 dollars E.-U par an. L'impact sur son beau-fils risque d'être moindre en raison d'une formation de qualité inférieure, mais la formation d'une personne a amélioré incontestablement les perspectives de subsistance de deux personnes. Le coût par personne est donc plus faible que les 1000 dollars E.-U initialement calculés.

Les avantages indirects peuvent donc être importants et pourraient, dans certains cas, justifier les coûts d'un programme. Ces effets en cascade devraient être évalués lors de la collecte finale, en demandant aux participants si d'autres personnes ont reçu une formation ou bénéficié du programme.

Les effets prix et quantité

Même si les interventions peuvent cibler certains marchés ou population locales, celles-ci peuvent avoir des effets sur l'ensemble de l'économie, communément appelés : effets « d'équilibre général ». En reprenant notre exemple, si consécutivement à la formation en menuiserie des menuisiers qualifiés supplémentaires (hors programme) sont apparus dans l'économie locale, leur concurrence peut entraîner une diminution des prix pour le consommateur. Il est également possible que le programme ait des retombées négatives sur l'économie. L'installation d'artisans tailleurs supplémentaires dans une zone où cette profession est déjà largement représentée, par exemple, pourrait conduire à une chute drastique des prix, et à la cessation d'activité de certains artisans. Ce type de situation mettrait un sérieux bémol aux objectifs du programme.

D'autres effets indésirables du programme peuvent porter préjudice aux non-participants. Par exemple, si les participants à l'intervention obtiennent un avantage concurrentiel par rapport aux autres sur le marché du travail, certains jeunes ne trouveront pas d'emploi, alors qu'ils auraient pu en trouver en l'absence de programme. Ces effets sont communément appelés les effets de déplacement.

Dans la plupart des programmes, les effets prix et quantité sont susceptibles d'être minimes, et ne nécessitent pas la collecte de données. Cependant, certains dirigeants de programmes importants peuvent souhaiter évaluer ces effets aidés d'un expert en évaluation. Il est envisageable de rendre aléatoire l'intervention au niveau des collectivités ou du district.

Points essentiels

1. En fonction des objectifs d'apprentissage de l'organisation et du programme, il est utile de déterminer si une évaluation d'impact peut être programmée pour mesurer d'autres paramètres que l'impact moyen du programme. Ces nouvelles supplémentaires peuvent se poser en matière d'hétérogénéité, d'horizon temporel, d'effets d'entraînement, ou d'efficacité relative des différentes options liées à la conception d'un programme.

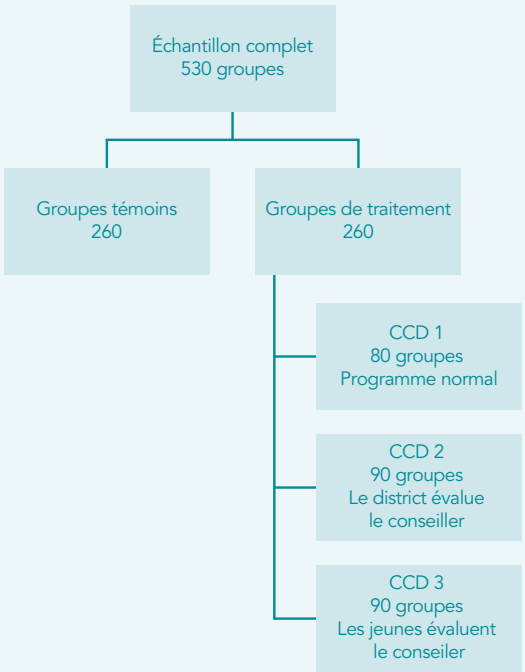
2. Il est vivement recommandé d'introduire des éléments de recherche qualitative dans une évaluation d'impact. Utiliser des méthodes mixtes permet une compréhension plus complète et plus nuancée de l'impact ou de l'absence d'impact d'un programme.
3. Les informations sur l'impact d'un programme peuvent être d'une utilité limitée tant que l'on ne connaît pas les coûts de conception et de mise en œuvre de l'intervention. Le développement du programme dépendra de ce type d'information. Il est donc souhaitable de compléter une évaluation d'impact avec une analyse coût-efficacité ou une analyse coûts-bénéfices.

Étude de cas du NUSAF : Améliorer la pertinence d'une évaluation d'impact

Les modèles transversaux (CCD)

En plus d'évaluer les effets globaux du Youth Opportunities Program, l'évaluation d'impact a été mise à profit pour tester une intervention pilote complémentaire sur une conception novatrice de programme. Des témoignages rapportés de précédents cycles de financement de programmes suggèrent que la gestion de la qualité, de la planification, des services de vulgarisation fournis par le district et les facilitateurs communautaires sont hautement déterminants pour la réussite du groupe de jeunes. L'évaluation d'impact a donc voulu estimer l'efficacité d'une dépense supplémentaire au vu d'engager un conseiller en vulgarisation, choisi par le groupe de jeunes lui-même.

Les groupes de traitement ont été randomisés pour participer au modèle transversal. Les projets financés ont été assignés aléatoirement à l'un des trois groupes. Le groupe de traitement 1 est traité normalement, sans aucune intervention supplémentaire. Pour le groupe de traitement 2, les officiers du district évaluent le conseiller en vulgarisation. Le groupe de traitement 3 doté de ressources supplémentaires a demandé d'évaluer le conseiller en vulgarisation lui même.



(suite page suivante)

Étude de cas du NUSAF : Améliorer la pertinence d'une évaluation d'impact (suite)

Méthodes mixtes

L'évaluation du Youth Opportunities Program a eu recours à des questionnaires quantitatifs et qualitatifs. Le questionnaire quantitatif a été soumis à environ 2 600 jeunes, et le questionnaire qualitatif à environ 100 jeunes. Les questions qualitatives portaient sur les sujets suivants :

1. Qualité de la coopération et de la dynamique de groupe comprenant le processus de formation du groupe ; structures et leadership du groupe ; processus de prise de décisions au sein du groupe ; activités passées, présentes et futures du groupe ; avantages et inconvénients du travail de groupe ; motivations personnelles du choix de travailler en groupe malgré les inconvénients.
2. Allocation de fonds du NUSAF comprenant le processus d'allocation au groupe, les priorités de financement du groupe par rapport aux plans initiaux du projet, ainsi que les déviations et autres usages non-officiels des fonds.
3. Formation comprenant le processus de choix de la formation professionnelle du groupe, assurance dans la mise en pratique des compétences acquises, avantages et inconvénients de la mise en pratique des compétences acquises en tant que stratégie de subsistance.
4. Stratégies de subsistance comprenant l'autonomisation après la formation professionnelle ; risques, succès et échecs associés aux nouvelles stratégies de subsistance ; raisons expliquant le succès ou l'échec ; stratégies de subsistance alternatives, et autres stratégies permettant de faire face aux risques et aux chocs.
5. Responsabilisation et participation à la vie communautaire comprenant le sens d'appartenance à la communauté, participation à la vie civile, relations entre hommes et femmes, soutien social, barrières sociales, et relations de voisinage.

Les données quantitatives ne peuvent à elles seules faire ressortir toute la richesse d'un programme. Les réponses à ces questions seront utilisées dans le but de mieux comprendre comment le NUSAF et le programme ont changé la vie des participants.

Source : Blattman, Martinez et Fiala (2011).

À lire

Bamberger, M., Rao, V., et Woolcock, M. 2010. *Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development.* Policy Research Working Paper 5245, Washington, DC: The World Bank. http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2010/03/23/000158349_20100323100628/Rendered/PDF/WPS5245.pdf

Bamberger, M., Rugh, J., et Mabry, L. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints.* Thousand Oaks: Sage Publications. (See chapter 13.) http://realworldevaluation.org/

Cellini, S. R., et Kee, J. E. 2010. "Cost-Effectiveness and Cost-Benefit Analysis." In: Wholey, J., Hatry, H. P., and Newcomer, K. E., eds. *Handbook of Practical Program Evaluation,* 3rd ed. San Francisco: Jossey-Bass. http://home.gwu.edu/~scellini/CelliniKee21.pdf

Knowles, J., et Behrman, J. 2005. *A Practical Guide to Economic Analysis of Youth Projects. HNP Discussion Paper.* Washington, DC: The World Bank. http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1095698140167/KnowlesPracticalGuideFinal.pdf

Notes

Ouvrages cités

Almeida, R., et Galasso, E. 2008. *Jumpstarting self-employment? Evidence from welfare participants in Argentina.* Washington, D.C. : La Banque mondiale. http://ftp.iza.org/dp2902.pdf

Attanasio, O., Kugler, A., et Meghir, C. 2009. "Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial." Discussion Paper No. 4251. Bonn : IZA. http://ftp.iza.org/dp4251.pdf

Baker, J. 2000. *Evaluation de l'impact des projets de Développement sur la pauvreté : Manuel à l'attention des Praticiens.* Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/03/29/000090341_20060329094344/Rendered/PDF/207450FRENCH0iehandbook1fr.pdf

Bamberger, M., Rao, V., et Woolcock, M. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper No. 5245. Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2010/03/23/000158349_20100323100628/Rendered/PDF/WPS5245.pdf

Banque asiatique de développement. 2007. *Handbook on Social Analysis, A Working Document.* Manille : BAsD. http://www.adb.org/Documents/Handbooks/social-analysis/Appendixes.pdf

Banque mondiale. 2002. *A Sourcebook for Poverty Reduction Strategies.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTPRS1/Resources/383606-1205334112622/4943_annex_c.pdf

———. 2006. *Rapport sur le développement dans le monde 2007 : Le développement et la prochaine génération.* Washington, D.C. : La Banque mondiale.http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/09/13/000112742_20060913111024/Rendered/PDF/359990WDR0complete.pdf

———. 2007a. *Data for Impact Evaluation.* Doing Impact Evaluation No. 6. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_06.pdf

———. 2007b. *Impact Evaluation for Microfinance: Review of Methodological Issues.* Doing Impact Evaluation No. 7. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_07.pdf

———. 2007c. "Miles to Go: A Quest for an Operational Labor Market Paradigm for Developing Countries." Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTLM/Resources/390041-1212776476091/5078455-1267646113835/MILESThequestoperationalLMparadigm_Jan212008.pdf

———. 2008. Project Paper for a project in the amount of USD 2,793,706 to the Republic of Liberia for a proposed Economic Empowerment of Adolescent Girls Project, Report No. 44985-LR. Washington, D.C. : La Banque mondiale.

———. 2009. *Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/EXTEVACAPDEV/Resources/4585672-1251461875432/inst_ie_framework_me.pdf

———. 2011. "Evaluation of the Use of Entertainment Education to Improve Financial Capability in South Africa." Note de synthèse (Février). Washington, D.C. : La Banque mondiale.

Beauvy-Sany, M., Conklin, S., Hershkowitz, A., Rajkotia, R., et, Berg, C. 2009. *Directives et expériences visant à inclure les jeunes dans les évaluations de marchés pour le renforcement des programmes de développement de la force de travail axée sur les jeunes.* Washington, D.C. : Le Réseau SEEP. http://www.seepnetwork.org/Resources/YouthPLP_Assessments.pdf

Bertrand, A., Beauvy-Sany, M., Cilimkovic, S., Conklin, S., et Jahic, S. 2009. *Monitoring and Evaluation for Youth Workforce Development Projects.* Washington, D.C. : Le Réseau SEEP. http://seep-network.org/Resources/YouthPLP_MonitoringEval.pdf

Betcherman, G., Godfrey, M., Puerto, S., Rother, F., et Stavreska, A. 2007. « Inventaire mondial des interventions d'appui à l'emploi des jeunes : Rapport de synthèse. » SP Discussion Paper No. 0715. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/SOCIALPROTECTION/Resources/SP-Discussion-papers/Labor-Market-DP/0715_Fr.pdf

Blattman, C., Martinez, S., et Fiala, N. 2009. *Impact Evaluation Baseline Report: Northern Uganda Social Action Fund Youth Opportunities Program.* Washington, D.C.: La Banque mondiale. http://www.iza.org/conference_files/ELMPDC2009/martinez_s4899.pdf

Bose, R. 2010. "A Checklist for the Reporting of Randomized Control Trials of Social and Economic Policy Interventions in Developing countries: CEDE Version 1.0." Working Paper No.6. New Delhi : Initiative internationale pour l'évaluation d'impact (3ie). http://www.3ieimpact.org/admin/pdfs_papers/working paper 6.pdf

Brady, M., Salem, A., et Zibani, N. 2007. "Bringing New Opportunities to Adolescent Girls in Socially Conservative settings: The Ishraq Program in Rural Upper Egypt." *Transitions to Adulthood*, Brief No. 12. New York : Conseil de population. http://www.popcouncil.org/pdfs/IshraqFullReport.pdf

Brener, N. D., Billy, J. O .G., et Grady, W.R. 2003. "Assessment of Factors Affecting the Validity of Self-Reported Health-Risk Behavior Among Adolescents: Evidence From the Scientific Literature." *Journal of Adolescent Health 33: 436–457*. http://www.cdc.gov/HealthyYouth/yrbs/pdf/validity.pdf

Bronfenbrenner, U. 1979. *The Ecology of Human Development: Experiments by Nature and Design.* Cambridge, MA : Harvard University Press. http://books.google.com/books?id=OCmbzWka6xUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Bruhn, M., et Zia, B. 2011. "Stimulating Managerial Capital in Emerging Markets: The Impact of Business and Financial Literacy for Young Entrepreneurs." Policy Research Working Paper No. 5642. Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2011/04/27/000158349_20110427082512/Rendered/PDF/WPS5642.pdf

Card, D., Kluve, J., et Weber, R. 2009. "Active Labor Market Policy Evaluations: A Meta-Analysis." IZA Discussion Paper No. 4002. Bonn. http://ftp.iza.org/dp4002.pdf

Cooley, L. 1989. "The Logical Framework: Program Design for Program Results." *The Entrepreneurial Economy Review 8(1): 8–15*. http://evaluation.zunia.org/post/the-logical-framework-program-design-for-program-results/

Creswell, John. 2008. *Research Design: Qualitative, Quantitative and Mixed Methods Approaches.* Thousand Oaks, CA : Sage. http://books.google.com/books?id=bttwENORfhgC&printsec=frontcover&source=gbs_ge_summary_r&cad=0

Cunningham, W., Sanchez-Puerta, M. L., et Wuermli, A. 2010. Active Labor Market Policies for Youth: A Framework to Guide Youth Employment Interventions. Employment Policy Primer No. 16. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTLM/214578-1103128720951/22795057/EPPNoteNo16_Eng.pdf

Davies, R. J., et Dart, J. 2005. *La Technique du 'Changement le Plus Significatif' (CPS) : Guide de l'Utilisateur*. http://www.mande.co.uk/docs/MSCtranslations/MSC%20Guide%20FRENCH_Oct%202007.doc

Development Marketplace. 2008. *Development Marketplace Grantee Toolkit: Project Design, Monitoring and Evaluation for Small Innovative Projects.* Washington D.C. : La Banque mondiale. http://siteresources.worldbank.org/DEVMARKETPLACE/Resources/DM_Grantee_Toolkit_Final.pdf

DFID (Ministère du développement international). 1999. "Cadre" Notes d'information sur les moyens de subsistance durables, Section 2. Oxford : DFID. http://community.eldis.org/.59c21877/FR-GS2.pdf

Duflo, E., Glennerster, R., et Kremer, M. 2006. "Using Randomization in Development Economics Research: A Toolkit." Bureau for Research and Economic Analysis of Development (BREAD), Working Paper No. 136. http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf

Fiszbein, A., et Schady, N. 2009. *Transferts monétaires conditionnels : Réduire la pauvreté actuelle et future.* Policy Research Report. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTCCT/Resources/5757608-1234228266004/CCTS-Overview-French.pdf

GAO (U.S. Government Accountability Office). 1991. *Designing Evaluations.* Washington, D.C. : GAO. http://www.gao.gov/special.pubs/10_1_4.pdf

Gawler, M. 2005. *Useful Tools for Engaging Young People in Participatory Evaluation.* Genève : Bureau régional de l'UNICEF pour l'Europe centrale et orientale et la Communauté des états indépendants. http://www.artemis-services.com/downloads/tools-for-participatory-evaluation.pdf

Gertler, P., Martinez, S., Premand, P., Rawlings, L., et Vermeersch, C. 2011. *L'évaluation d'impact en pratique.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTHDOFFICE/Resources/IEP_Ebook_FRENCH.pdf

Gosparini, P., Russo, L., Sirtori, M., et Valmarana, C. 2004. *The Monitoring and Evaluation Manual of the NGOs of the Forum Solint.* Rome (Italie) : CISP, COSV, COOPI, Intersos, Movimondo, et DRN. http://www.cosv.org/echotrain/materiale/0B_ITA/ECHOTrain_Documenti/ECHOTrain_Documenti_Manuali/ECHOTrain_Documenti_Manuali_SOLINT/Manuale M&E- Solint.pdf

Hempel, K. 2006. Erfolgskontrolle in der deutschen Entwicklungszusammenarbeit: Zentrale Herausforderungen unter besonderer Berücksichtigung der Transparenz gegenüber der Öffentlichkeit. Stuttgart: ibidem-Verlag.

Ibarrarán, P., et Rosas Shady, D. 2009. "Evaluating the Impact of Job Training Programmes in Latin America: Evidence from IDB-Funded Operations." *Journal of Development Effectiveness* 1(2): 195–216. http://www.iza.org/conference_files/ELMPDC2009/ibarraran_p4263.pdf

Imas, L., et Rist, R. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations.* Washington, D.C. : La Banque mondiale. http://books.google.com/books?id=NEsg-BtinIsC&printsec=frontcover&source=gbs_ge_summary_r&cad=0 - v=onepage&q&f=false

Innovations for Poverty Action. 2010. *IPA IRB Submission Guidelines and Application.* Document interne. New Haven, CT : IPA.

Jaramillo, M., et Parodi, S. 2003. Jóvenes Emprendedores. Lima, Pérou : Instituto Apoyo. http://www.grade.org.pe/download/pubs/MJ-SP-J%C3%B3venes%20emprendedores.pdf

Jones, N., Jones, H., Steer, L., et Datta, A. 2009. "Improving impact evaluation production and use." Working Paper No. 3000. Londres : Overseas Development Institute. http://www.odi.org.uk/resources/download/3177.pdf

J-PAL (Laboratoire d'Action contre la Pauvreté Abdul Latif Jameel). 2011. "Cost-benefit/Effectiveness/Comparison Analysis." Site Internet J-PAL. http://www.povertyactionlab.org/methodology/what-evaluation/cost-benefit/effectiveness/comparison-analyses. Au 31 mai 2011.

Karlan, D. 2009. "Thoughts on Randomised Trials for Evaluation of Development: Presentation to the Cairo Evaluation Clinic." *The Journal of Development Effectiveness* 1(3): 237–242. http://www.3ieimpact.org/admin/pdfs_papers/50.pdf

Kellogg (Fondation W. K. Kellogg). 1998. *Evaluation Handbook*. Battle Creek, MI : Fondation W. K. Kellogg. http://www.wkkf.org/knowledge-center/resources/2010/W-K-Kellogg-Foundation-Evaluation-Handbook.aspx

Kellogg (Fondation W. K. Kellogg). 2004. *Logic Model Development Guide*. Battle Creek, MI : Fondation W. K. Kellogg. http://www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx

Khandker, S., Koolwal, G., et Samad, H. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101Official0Use0Only1.pdf

Klinger, B., et Schuendeln, M. 2007. "Can Entrepreneurial Activity be Taught? Quasi-Experimental Evidence from Central America." Working Paper No. 153. Cambridge, MA : Harvard University Center for International Development. http://www.hks.harvard.edu/var/ezp_site/storage/fckeditor/file/pdfs/centers-programs/centers/cid/publications/faculty/wp/153.pdf

Kusek, J. Z., et Rist, R. C. 2004. *Dix étapes pour mettre en place un système de suivi et d'évaluation axé sur les résultats : Un guide pour les praticiens du développement.* http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/11/16/000310607_20061116133234/Rendered/PDF/296720FRENCH0T1d1OFFICIAL0USE0ONLY1.pdf

Leeuw, F., et Vaessen, J. 2009. *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation.* Washington, D.C. : Réseau de réseaux pour l'évaluation d'impact (NONIE). http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf

Muzi, S. 2011. "Summary Notes on Impact Evaluations in the Active Labor Market/Youth Employment Programs Cluster Supported by SIEF." Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1239047988859/5995659-1239048041095/cf4_zMisc_ALM_SummaryNotes_SIEF-IE-on-ALMP_Jan5_PA_110121.pdf

NIH (National Institutes of Health). 2008. *Protecting Human Research Participants.* Bethesda : NIH Office of Extramural Research. http://phrp.nihtraining.com/users/PHRP.pdf

NORC (National Opinion Research Center). 2007. *Operational Plan for the Baseline Survey of "Mi Primer Empleo" Youth Employment Program, Honduras.* Chicago : NORC.

OCDE (Organisation de coopération et de développement économique). 1991. *Principes pour l'évaluation de l'aide au développement.* Paris : Comité d'aide au développement. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=OCDE/GD(91)208&docLanguage=Fr

OCDE. (n.d.). DAC Criteria for Evaluating Development Assistance. http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html

Osborn, D., et Gaebler, T. 1992. *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Pubic Sector*. Reading, MA : Addison-Wesley.

Penrose-Buckley, C. 2007. "Annex 2: Rapid Assessment of Markets and Producers." In *Producer Organisations: A Guide to Developing Collective Rural Enterprises.* Oxford : Oxfam RU http://www.oxfam.org.uk/resources/downloads/produorgs_book.pdf

Powers, J. L., et Tiffany, J. S. 2006. "Engaging Youth in Participatory Research and Evaluation." *Journal of Public Health Management and Practice*, novembre (suppl.), S79-S87. http://www.health.state.ny.us/community/youth/development/docs/jphmp_s079-s087.pdf

Ravallion, M. 2008. *Evaluating Anti-Poverty Programs. Handbook of Development Economics*, Vol. 4, Schultz, T. P., et Strauss, J. (éd.). Amsterdam : Elsevier. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1130267506458/Evaluating_Antipoverty_Programs.pdf

Rawlings, L., et Rubio, G. 2005. "Evaluating the Impacts of Conditional Cash Transfer Programs." *World Bank Research Observer* 20(1): 29–55. http://www.crin.org/docs/Evaluating the Impact of Cash Transfer Programs.pdf

RMC (RMC Research Corporation). 2009. *"Our Nation" Evaluation*. Denver, CO : RMC. http://www.myja.org/programs/evaluation/reports/our_nation.pdf

Roberts, B., Caspi, A., et Moffitt, T. 2003. "Work Experiences and Personality Development in Young Adulthood." *Journal of Personality and Social Psychology* 84(3): 582–593. http://www.mendeley.com/research/work-experiences-and-personality-development-in-young-adulthood-1/

Robins, R., Fraley, R., Roberts, B., et Trzesniewski, K. 2001. "A Longitudinal Study of Personality Change in Young Adulthood." *Journal of Personality* 69(4): 617–640. http://internal.psychology.illinois.edu/~broberts/Robins,%20et%20al,%202001.pdf

Rubio, G. 2011. "The Design and Implementation of a Menu of Evaluations." PREM Notes No. 6, The Nuts and Bolts of M and E Systems. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTPOVERTY/Resources/335642-1276521901256/premnoteME6.pdf

Rossiasco, P., O'Neil, F., Richardson, A., et Francis, P. 2010. "Youth and Employment In Post-Conflict Countries: The Psycho-Social Dimension." *Child and Youth Development Notes* (4)2. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTLM/Resources/390041-1319047943696/CYDN_No2_Psychosocial_Employment.pdf

Sabo Flores, K. 2008. *Youth Participatory Evaluation: Strategies for Engaging Young People.* San Francisco : Jossey-Bass. http://www.josseybass.com/WileyCDA/WileyTitle/productCd-0787983926.html

Savedoff , W. D., Levine, R., et Birdsall, N. 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation.* Washington, D.C. : Center for Global Development. http://www.cgdev.org/content/publications/detail/7973

Shapiro, J. 2003. "Monitoring and Evaluation." Washington, D.C. : CIVICUS. http://www.civicus.org/new/media/Monitoring and Evaluation.doc

Taylor-Powell, E. 2005. "Logic Models: A Framework for Program Planning and Evaluation." University of Wisconsin–Extension, Program Development and Evaluation, 2005. Diapositives 15–16. http://www.uwex.edu/ces/pdande/evaluation/pdf/nutritionconf05.pdf. Au 2 juin 2011.

UNICEF (Fonds des Nations Unies pour l'enfance). 1991. "A UNICEF Guide for Monitoring and Evaluation: Making a Difference?" New York : UNICEF. http://preval.org/documentos/00473.pdf

Trochim, W. M. K. 2006. "The Nonquivalent Groups Design." Web Center for Social Research Methods, Research Methods Knowledge Base. http://www.socialresearchmethods.net/kb/quasnegd.php. Au 2 juin 2011.

Bibliographie

Documentation sur le développement et l'emploi des jeunes

Banque mondiale 2006. *Rapport sur le développement dans le monde 2007 : Le développement et la prochaine génération.* Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/09/13/000112742_20060913111024/Rendered/PDF/359990WDR0complete.pdf

Bidwell, K., Galbraith, C., et al. 2008. *Market Assessment Toolkit for Vocational Training Providers and Youth.* New York : Women's Commission for Refugee Women and Children et Columbia University School of International and Public Affairs. http://www.womensrefugeecommission.org/docs/ug_ysl_toolkit.pdf

Bronfenbrenner, U. 1979. *The Ecology of Human Development: Experiments by Nature and Design.* Cambridge, MA : Harvard University Press. http://books.google.com/books?id=OCmbzWka6xUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0 - v=onepage&q&f=false

Bureau international du Travail. 2010. *Global Trends for Youth: Special Issues on the Impact of the Global Economic Crisis on Youth.* Genève : BIT. http://www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_elm/---trends/documents/publication/wcms_143349.pdf

Cunningham, W., Cohan, L., Naudeau, S., et McGinnis, L. 2008. *Supporting Youth at Risk: A Policy Toolkit for Middle-Income Countries.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTCY/Resources/395766-1187899515414/SupportingYouthAtRisk.pdf

Cunningham, W., Sanchez-Puerta, M. L., et Wuermli, A. 2010. *Active Labor Market Policies for Youth: A Framework to Guide Youth Employment Interventions.* Employment Policy Primer No. 16. Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTLM/214578-1103128720951/22795057/EPPNoteNo16_Eng.pdf

James-Wilson, D. 2008. *Youth Livelihoods Development Program Guide*. Washington, D.C. : EQUIP3. http://www.equip123.net/docs/e3-LivelihoodsGuide.pdf

Making Cents International. *State of the Field in Youth Enterprise Development and Livelihoods Development series.* Washington, D.C. : MCI. http://www.youtheconomicopportunities.org/media.asp

Van Adams, A. 2007. *The Role of Youth Skills in the Transition to Work: A Global Review.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTCY/Resources/395766-1187899515414/RoleofYouthSkills.pdf

Van der Geest, K. 2010. *Rural Youth Employment in Developing Countries: A Global View.* Rome : FAO. http://www.fao.org/docrep/012/al414e/al414e00.pdf

Women's Refugee Commission. 2009. *Building Livelihoods: A Field Manual for Practitioners in Humanitarian Settings.* New York : WRC. http://www.womensrefugeecommission.org/docs/livelihoods_manual.pdf

Documentation sur le suivi et l'évaluation

Banque mondiale 2007. Doing Impact Evaluation. Série de quatorze rapports associés. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:22268536~pagePK:210058~piPK:210062~theSitePK:384329,00.html

Bamberger, M., Rugh, J., et Mabry, L. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints.* Thousand Oaks : Sage. http://www.realworldevaluation.org

Bertrand, A., Beauvy-Sany, M., Cilimkovic, S., Conklin, S., et Jahic, S. 2009. Monitoring and Evaluation for Youth Workforce Development Projects. Washington D.C.: Le Réseau SEEP. http://seepnetwork.org/Resources/YouthPLP_MonitoringEval.pdf

Catley, A., Burns, J., Abebe, D., et Suji, O. 2010. *Évaluation d'impact participative : guide à destination des praticiens.* Medford, MA : Feinstein International Center. https://wikis.uit.tufts.edu/confluence/download/attachments/19924843/Part_Impact_10_21_08V2.pdf?version=1&modificationDate=1225200269000 (version anglaise) http://reliefweb.int/sites/reliefweb.int/files/resources/5D6A92EEAA3D982B432575C400445332-Tufts_Mai2009.pdf (version française)

Duflo, E., Glennerster, R., et Kremer, M. 2006. "Using Randomization in Development Economics Research: A Toolkit.", Bureau for Research and Economic Analysis of Development (BREAD) Working Paper No. 136. http://www.povertyactionlab.org/sites/default/files/documents/Using%20Randomization%20in%20Development%20Economics.pdf

Gertler, P. Martinez, S., Premand, P., Rawlings, L., et Vermeersch, C. 2011. *L'évaluation d'impact en pratique.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/INTHDOFFICE/Resources/IEP_Ebook_FRENCH.pdf

Imas, L., et Rist, R. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations.* Washington, D.C. : La Banque mondiale. http://books.google.com/books?id=NEsg-BtinIsC&printsec=frontcover&source=gbs_ge_summary_r&cad=0

Karlan, D., et Appel, J. 2011. *More than Good Intentions: How a New Economics Is Helping to Solve Global Poverty.* Boston, MA : Dutton. http://books.google.com/books?id=JOMDsm5Gn9kC&printsec=frontcover&source=gbs_ge_summary_r&cad=0 - v=onepage&q&f=false

Khandker, S., Koolwal, G., et Samad, H. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices.* Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2009/12/10/000333037_20091210014322/Rendered/PDF/520990PUB0EPI1101Official0Use0Only1.pdf

Kluve, J. 2011. *Measuring Employment Effects of Technical Cooperation Interventions: Some Methodological Guidelines.* Second revised edition. Eschborn : GIZ http://www2.gtz.de/wbf/4tDx9kw63gma/Methodenleitfaden.pdf

Knowles, J., et Behrman, J. 2005. *A Practical Guide to Economic Analysis of Youth Projects. HNP Discussion Paper.* Washington, D.C. : La Banque mondiale. http://siteresources.worldbank.org/HEALTHNUTRITIONANDPOPULATION/Resources/281627-1095698140167/KnowlesPracticalGuideFinal.pdf

Kusek, J. Z., et Rist, R. C. 2004. *Dix étapes pour mettre en place un système de suivi et d'évaluation axé sur les résultats : Un guide pour les praticiens du développement.* Washington, D.C. : La Banque mondiale. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2006/11/16/000310607_20061116133234/Rendered/PDF/296720FRENCH0T1d1OFFICIAL0USE0ONLY1.pdf

Leeuw, F., et Vaessen, J. 2009. *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation.* Washington, D.C. : Réseau de réseaux pour l'évaluation d'impact (NONIE). http://siteresources.worldbank.org/EXTOED/Resources/nonie_guidance.pdf

Powell, E., Jones, L., et Henert, E. 2003. *Enhancing Program Performance with Logic Models.* University of Wisconsin–Extension, Program Development and Evaluation. http://www1.uwex.edu/ces/lmcourse/

Ravallion, M. 2008. *Evaluating Anti-Poverty Programs. Handbook of Development Economics,* Vol. 4, Schultz, T. P., et Strauss, J. (éd.). Amsterdam : Elsevier. http://siteresources.worldbank.org/INTISPMA/Resources/383704-1130267506458/Evaluating_Antipoverty_Programs.pdf

Wholey, J., Hatry, H. P., et Newcomer, K. E., éd. 2010. *Handbook of Practical Program Evaluation,* 3e éd. San Francisco : Jossey-Bass. http://books.google.com/books?id=zntNhoO6gCUC&printsec=frontcover&source=gbs_ge_summary_r&cad=0 - v=onepage&q&f=false

Bases de données de recherche

EconPapers http://econpapers.repec.org/

Education Resources Information Center (ERIC) http://www.eric.ed.gov/

IDEAS http://ideas.repec.org/

IZA Discussion Papers http://www.iza.org/en/webcontent/publications/papers

J-STOR http://www.jstor.org/

National Bureau of Economic Research (NBER) http://www.nber.org/

PsychInfo http://www.apa.org/pubs/databases/psycinfo/index.aspx

Social Science Research Network http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm

Journaux académiques

Journal of Adolescence
http://www.elsevier.com/wps/find/journaldescription.cws_home/622849/description#description

Journal of Early Adolescence
http://jea.sagepub.com/

Journal of Research on Adolescence
http://www.wiley.com/bw/journal.asp?ref=1050-8392

Journal of Youth and Adolescence
http://www.springer.com/psychology/child+%26+school+psychology/journal/10964

Journal of Youth Development
www.nae4ha.org/directory/jyd/index.html

Journal of Youth Studies
http://taylorandfrancis.co.uk/journals/titles/13676261.asp

New Directions for Child and Adolescent Development
http://www3.interscience.wiley.com/journal/85511342/home

New Directions for Youth Development
http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1537-5781/homepage/ProductInformation.html

The Future of Children
http://futureofchildren.org/

Vulnerable Children and Youth Studies
http://www.informaworld.com/smpp/title~db=all~content=t724921266

Young: Nordic Journal of Youth Research
http://you.sagepub.com

Bases de données existantes et permanentes sur les évaluations d'impact

Laboratoire d'Action contre la Pauvreté Abdul Latif Jameel (J-PAL)
http://www.povertyactionlab.org/search/apachesolr_search?filters=type:evaluation

Initiative internationale pour l'évaluation d'impact (3ie)
http://www.3ieimpact.org/database_of_impact_evaluations.html

Innovations for Poverty Action (IPA)
http://www.poverty-action.org/project-evaluations/search

World Bank DIME Initiative
http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDEVIMPEVAINI/0,,contentMDK:21553788~pagePK:64168445~piPK:64168309~theSitePK:3998212,00.html

Youth Employment Inventory (Inventaire sur l'emploi des jeunes)
http://www.youth-employment-inventory.org/

Blogues

Aid Watch
http://aidwatchers.com/

Chris Blattman
http://chrisblattman.com/

Development that Works (IDB)
http://blogs.cross-cutting.org/desarrolloefectivo_en/

Find What Works
http://findwhatworks.wordpress.com/

Good Intentions Are Not Enough
http://goodintents.org/

Innovations for Poverty Action
http://www.poverty-action.org/blog

Larry Dershem, Save the Children
http://designmonitoringevaluation.blogspot.com/

Monitoring and Evaluation News
http://mande.co.uk/

World Bank Development Impact
http://blogs.worldbank.org/impactevaluations/

Renforcement des capacités

Cliniques d'évaluation du YEN
http://yenclinic.groupsite.org

Cours du Centre international de formation de l'OIT sur le suivi et l'évaluation de l'impact des programmes pour l'emploi des jeunes
http://www.itcilo.org

International Program for Development Evaluation Training (IPDET)
http://www.ipdet.org

J-PAL Course on Impact Evaluation
http://www.povertyactionlab.org/course

Il existe un grand choix de formations en auto-apprentissage, en ligne ou sur place en matière de suivi et d'évaluation (au-delà de l'évaluation d'impact) sur le site Internet de My M&E,
http://www.mymande.org/index.php?q=training_search&x=admin

World Bank Impact Evaluation Workshops
http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:21754074~menuPK:384336~pagePK:148956~piPK:216618~theSitePK:384329,00.html

Sites Internet

Evalsed: Ressource pour l'évaluation du développement socio-économique
http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/

My M&E
http://www.mymande.org/

Web Center for Social Research Methods
http://www.socialresearchmethods.net/kb/design.php

ANNEXE 1. Exemples d'indicateurs pour les évaluations de la jeunesse

Statut des jeunes	
Catégorie	**Indicateurs sélectionnés**
Pauvreté	• Nombre et pourcentage de jeunes vivant avec moins de 2 USD par jour • Nombre et pourcentage de jeunes souffrant de malnutrition • Nombre et pourcentage de jeunes vivant dans des foyers sans infrastructure de base
Niveau d'éducation	• Taux d'alphabétisation des jeunes • Taux net de scolarisation dans l'enseignement primaire • Taux net de scolarisation dans l'enseignement secondaire • Taux brut de scolarisation dans l'enseignement supérieur • Ratio de scolarisation féminine par rapport à la scolarisation masculine, selon le niveau d'éducation • Niveau de scolarité à l'âge x
Situation d'emploi	• Pourcentage de jeunes employés, au chômage et sous-employés • Taux de jeunes faisant partie de la population active • Taux de chômage des jeunes par rapport à celui des adultes • Nombre et pourcentage de jeunes non scolarisés, en formation ou en emploi • Emploi des jeunes par secteur (ou par type de profession) • Emploi des jeunes selon le type de travail (salariés, indépendants, employeurs, travail à la maison ou travail non rémunéré) • Emploi des jeunes selon le type de contrat (formel ou informel, à plein temps ou à temps partiel, à durée indéterminée ou déterminée) • Revenu mensuel moyen • Temps nécessaire (moyenne) pour trouver un emploi • Disponibilité de stages et d'apprentissages • Préférences des jeunes concernant l'emploi (administration, secteur privé, indépendants, etc.) et attitudes concernant le travail
Santé	• Prépondérance du VIH/SIDA chez les jeunes (en %) • Nombre et pourcentage de jeunes ayant une bonne connaissance du VIH/SIDA et de la contraception • Nombre et pourcentage de jeunes qui utilisent actuellement une méthode contraceptive • Nombre et pourcentage de jeunes qui déclarent avoir eu des rapports sexuels à risque pendant l'année qui précède • Prépondérance de consommation de drogues (par exemple, cannabis, cocaïne, substances inhalées) • Prépondérance de la consommation de tabac et d'alcool chez les jeunes
Situation familiale	• Moyenne d'âge lors du premier rapport sexuel • Nombre ou pourcentage de femmes entre 15 et 19 ans qui ont des enfants ou qui sont enceintes • Pourcentage de femmes avec au moins un enfant, selon l'âge • Ratio de filles mariées avant 18 ans • Moyenne d'âge lors du premier mariage, selon le sexe
Antécédents criminels	• Nombre et pourcentage de jeunes qui ont commis ou qui ont été victimes de crimes et de violence • Nombre et pourcentage de jeunes qui font partie de bandes ou de groupes rebelles • Nombre et pourcentage de jeunes touchés par la violence domestique • Nombre et pourcentage de jeunes touchés par la violence sexuelle • Nombre et pourcentage de jeunes ayant affaire avec la justice

Exemples d'indicateurs pour les évaluations de la jeunesse (suite)

Citoyenneté	• Nombre et pourcentage de jeunes en âge de voter qui votent aux élections • Nombre et pourcentage de jeunes qui s'engagent dans des activités bénévoles • Nombre et pourcentage de jeunes qui font partie d'organisations de jeunes, groupes, etc. • Raisons pour lesquelles les jeunes ne participent pas à des activités civiques
Vulnérabilité	• Nombre et pourcentage de jeunes vivant dans la rue • Nombre et pourcentage de jeunes victimes de la prostitution et/ou de la traite des personnes • Nombre et pourcentage de jeunes touchés par les conflits • Nombre et pourcentage de jeunes souffrant de handicaps mentaux ou physiques • Quelles sont les vulnérabilités spécifiques aux jeunes hommes et celles spécifiques aux jeunes femmes ?
Perceptions et aspirations	• Nombre et pourcentage de jeunes satisfaits avec leur vie actuelle • Nombre et pourcentage de jeunes qui ont confiance en l'avenir • Nombre et pourcentage de jeunes qui souhaitent émigrer • Nombre et pourcentage de jeunes qui font confiance aux adultes, dirigeants communautaires, politiciens, institutions, etc. Pour ceux qui ne sont pas satisfaits, que voudraient-ils leur dire? • Quelles aspirations ont les jeunes en relation avec leur éducation ? Comment conçoivent-ils et quelles différences voient-ils entre la formation professionnelle et la formation universitaire et pourquoi? • Quelles sont les priorités, ambitions et opportunités des jeunes? • Quels sont les plus grands défis, risques et frustrations auxquels les jeunes doivent faire face?
Environnement direct	
Catégorie	**Indicateurs sélectionnés**
Famille	• Taille moyenne du ménage • Pourcentage de ménages monoparentaux • Caractéristiques de la situation professionnelle des parents (statut professionnel, temps de travail, etc.) • Moyenne d'heures que les parents consacrent par semaine à leurs enfants • Relations parents-enfants
École	• Nombre et pourcentage de jeunes ayant une école secondaire à x kilomètres • Type d'écoles et d'instituts de formation dans leur région • Qualité des écoles et des instituts de formation existant dans leur région • Prépondérance des frais de scolarité • Que pensent les professeurs des jeunes
Voisinage / quartier	• Statut socioéconomique et inégalités • Dynamique sociale dans la communauté (par ex. confiance, réseaux sociaux, etc.) • Prépondérance du crime et de la violence • Dynamique de groupes
Services adaptés aux jeunes	• Nombre de services de santé, culturels, sportifs ou religieux (publics et privés) • Proportion de jeunes satisfaits avec ces services • Motifs de satisfaction ou d'insatisfaction

Exemples d'indicateurs pour les évaluations de la jeunesse (suite)

Environnement local	
Catégorie	**Indicateurs sélectionnés**
Économie locale	• Proportion d'entreprises existantes par secteur/industrie • Salaire mensuel moyen par industrie (en USD) • Demande de main d'œuvre par industrie • Que pense l'employeur des jeunes • Lacunes au niveau des compétences tel qu'indiquées par les employeurs
Gouvernement local	• Allocations budgétaires par secteur (infrastructure, éducation, justice, programmes spécifiques pour les jeunes, etc.) • Initiatives publiques existantes destinées aux jeunes
Accès et utilisation des technologies et des médias	• Pourcentage de jeunes qui utilisent des téléphones portables • Pourcentage de jeunes ayant accès à internet et aux ordinateurs • Taux de pénétration de la télévision, radio, etc. • Comment la technologie et les médias sont-ils utilisés? Dans quel but sont-ils utilisés?
Accès et utilisation de services financiers	• Nombre et pourcentage de jeunes ayant un compte bancaire • Types d'épargne disponibles • Sources de crédit • Conditions permettant d'emprunter de l'argent auprès d'une banque locale (âge minimum, garanties, etc.)
Environnement sociétal	
Catégorie	**Indicateurs sélectionnés**
Démographie du pays	• Population par âge et sexe • Pourcentage de la population ayant moins de 25/29 ans • Pourcentage de la population ayant entre 12/15 et 24/29 ans • Ratio hommes / femmes dans la population • Pourcentage de population urbaine et population rurale • Pourcentage de population jeune par religion et ethnie • Pourcentage de population jeune par langue principale
Cadre juridique	• Âge de la majorité • Âge requis pour voter • Âge minimum légal pour le mariage • Protections et obligations légales par âge • Droit coutumier
Croyances et idéologies dominantes	• Organisation sociale et hiérarchies • Lignée familiale • Structures matrimoniales • Schémas d'héritage • Système de valeurs

Note : Pour obtenir des informations plus détaillées, la plupart des indicateurs mentionnés ci-dessus peuvent être ventilés par âge, sexe, niveau d'éducation, revenu du ménage, etc.

ANNEXE 2. Solutions pour diminuer les coûts ?

Pour réduire les coûts de l'évaluation d'impact, il faut souvent être un peu créatif. Cependant, cela risque d'avoir des effets sur la qualité de l'évaluation, il convient donc de considérer chaque aspect de la question avec précaution.

Réduire le coût des consultants. Il n'est pas toujours nécessaire d'engager un expert international pour diriger l'évaluation. Travailler avec un spécialiste *via* une organisation ou une université partenaire (pour laquelle le consultant pourra utiliser les résultats de l'évaluation pour une publication de recherche) permet de faire beaucoup d'économies. De même, les étudiants en master ou en doctorat de sciences économiques, santé publique ou autres sciences sociales ont très souvent les outils d'analyse nécessaires à une évaluation et la conception du suivi, ainsi que pour effectuer l'analyse des données. De plus, ils seront probablement contents de travailler sur un programme concret.

Réduire les coûts de la collecte de données. Lorsque l'on finalise le plan et la méthodologie d'évaluation, l'investigateur principal devra déterminer si, et jusqu'à quel point, les coûts relatifs à la collecte de données peuvent être réduits.

- « **Recycler** » **les données existantes.** Il se peut qu'il ne soit pas nécessaire de une collecter des données si on peut se servir d'enquêtes locales, régionales ou nationales qui ont déjà été conduites antérieurement ou qui sont en phase de planification. S'il existe déjà des données, l'évaluation d'impact peut se limiter à analyser ces données. De même, si une autre enquête couvrant la même population d'intérêt que la nôtre est prévue, l'évaluation peut ajouter toute une série de questions pertinentes pour notre programme.
- **Réduire la taille de l'échantillon.** Pour certains programmes, il est tentant de réduire le nombre de personnes à interviewer pour réduire les coûts. Malgré une diminution des coûts, cela augmente parallèlement la probabilité de conclure, à tort, que le programme n'a aucun impact, même si ce n'est pas le cas. Utiliser des échantillons réduits doit être fait avec précaution. L'expert en évaluation chargé de superviser l'évaluation d'impact indiquera la taille de l'échantillon minimale qui peut être utilisée dans un contexte donné.
- **Réduire la longueur du questionnaire d'enquête.** Les questionnaires d'enquête tendent à devenir longs, car les différentes parties prenantes suggèrent des questions (items) additionnelles intéressantes. Si on limite le questionnaire aux résultats d'intérêt principaux et si l'on exclut les questions qui ne sont pas directement liées aux objectifs principaux de l'intervention, on économisera du temps et de l'argent au cours de la collecte de données et de l'analyse.
- **Utiliser des instruments existants.** Il n'est pas toujours nécessaire de développer un nouveau questionnaire pour chaque évaluation. Si une organisation a une procédure d'évaluation type pour différentes interventions, on peut utiliser des questionnaires existants, internes ou externes à l'organisation (ou au moins une partie), qui ont déjà été validés. Il est souvent possible de créer un ensemble de questions et de modules de questionnaires qui pourront être utilisés lors d'évaluations suivant la même logique d'intervention et partageant les mêmes objectifs.
- **Utiliser le processus d'enregistrement du programme pour collecter des données de base.** Il est parfois possible de se servir du processus d'enregistrement du programme pour collecter les données individus s'inscrivant au programme. Cette information peut inclure les caractéristiques sociodémographiques de base ainsi que des informations relatives aux résultats d'intérêt qui, à leur tour, peuvent servir de base de données.
- **Collecte de données avec le personnel du programme ou du personnel local.** Faire la collecte de données à l'interne plutôt que de faire appel à un institut de sondage peut faire réaliser

des économies substantielles. On peut faire aussi des économies en engageant des professeurs (lorsque le programme est basé dans une école ou dans un centre de formation), des étudiants universitaires ou des membres de la communauté locale qui, à travail égal, sont prêts à recevoir une rétribution nettement inférieure à celle des employés d'un institut de sondage. Cela peut fonctionner pour les enquêtes brèves et peu complexes, mais pour les enquêtes à plus grande échelle, et plus approfondies cela présente des inconvénients. En effet, lorsque la collecte de données est plus complexe, le personnel du programme et le personnel local n'ont généralement ni l'expérience ni les compétences requises pour la mener. Les frais de supervision et de formation augmenteront, de même que le temps nécessaire pour finaliser la collecte. En cas de problèmes lors de la collecte, il se peut que les données obtenues ne soient pas très utiles, et une nouvelle collecte doit être mise en route, et cela peut s'avérer très coûteux. En outre, le fait que le personnel du programme soit chargé de la collecte de données soulève la question de leur neutralité, et risque de mettre en cause la fiabilité des résultats.

- **Utilisation de nouveaux outils de collecte de données.** Avec l'essor des nouvelles technologies, un nombre croissant de nouveaux outils technologiques peut être utilisé pour collecter les données. Par exemple, au lieu d'utiliser les questionnaires papier, on peut envisager de saisir les données directement sur place *via* un téléphone ou un ordinateur portables, rendant quasiment nuls le temps et le coût de saisie et de traitement de données.
- **Utilisation de questionnaires d'auto-évaluation ou des entretiens par téléphone.** Bien que cela permette de réduire les coûts de façon significative, se baser sur des questionnaires d'auto-évaluation ou des enquêtes par téléphone n'est pas une stratégie de collecte de données valable pour les interventions visant à améliorer les moyens de subsistance des jeunes dans les pays en voie de développement, notamment à cause du faible niveau d'instruction. Aussi faut-il utiliser ces techniques uniquement dans des circonstances très particulières, et après en avoir discuté longuement avec l'expert en évaluation.
- **Se concentrer sur des zones géographiques facilement accessibles.** On peut économiser de l'argent en limitant la collecte de données à des zones géographiques facilement accessibles. Si un programme est mis en œuvre dans plusieurs zones d'un même pays, l'évaluation peut se focaliser sur la capitale plutôt que sur les zones rurales reculées. Toutefois, cela signifie aussi que les résultats seront seulement représentatifs seulement pour le groupe de population et la zone géographique (le milieu urbain) traités lors de l'évaluation.
- **Réduction de l'enquête de base.** Il n'est pas toujours nécessaire de mener une enquête de base étendue, dans certains cas il n'est même pas nécessaire d'enquêter. Par exemple, lorsqu'un programme est bien élaboré et qu'il est accompagné d'un cadre de suivi solide, une enquête de base simple permettant de collecter rapidement des données sera suffisante pour comprendre la situation des participants et comparer le groupe de traitement et le groupe témoin. De même, pour les évaluations aléatoires, si la taille de l'échantillon est considérable, il y a peu de raisons d'être préoccupé par les différences entre le groupe de traitement et le groupe témoin. L'enquête de base n'est dans ce cas pas vraiment nécessaire.

  Toutefois, l'élimination d'une enquête de base comporte certains risques. En effet, elle permet d'avoir un premier contact avec les répondants, ce qui est souvent utile pour sélectionner la population qui sera suivie par les coordinateurs.

  Lors de l'enquête de base on doit s'assurer que les informations relatives à la localisation des participants sont très fiables ce qui permettra de les retrouver facilement lors de l'enquête de suivi. Les jeunes ont de fortes propensions à migrer. L'élimination de l'enquête de base pourrait donc augmenter considérablement les coûts liés à la recherche des participants une, deux ou

trois années plus tard. Sans cette enquête de base, il est également impossible d'utiliser la liste complète des caractéristiques des individus utilisée pour analyser les données. Les informations relatives au genre peuvent toujours être utilisées, mais on ne peut pas obtenir d'informations sur le bien-être au niveau de l'enquête de base, par exemple.

Recherche de financement externe. Le financement d'une évaluation d'impact ne doit pas être entièrement couvert par le budget du programme. Dans les faits, le financement est souvent complété par des sources externes, provenant de gouvernements, des Nations Unies, de banques de développement, de fondations, de philanthropes ou d'organismes de recherche et d'évaluation. L'importance croissante accordée à l'élaboration de politiques et de programmes basés sur les faits a augmenté la disponibilité de financements provenant de ces diverses sources. Quand une évaluation est potentiellement capable de combler une lacune importante dans un domaine de recherche qui intéresse le développement communautaire, comme c'est le cas pour la plupart des évaluations d'impact relatives aux moyens de subsistance des jeunes, les praticiens ont de bonnes raisons d'espérer qu'ils trouveront des financements externes.

Retarder la collecte de données de l'enquête de suivi. Parfois, la somme des fonds externes et internes n'est pas suffisante pour réunir le budget nécessaire qui couvrira l'ensemble de l'évaluation, impliquant probablement plusieurs cycles de collecte. Dans ce cas, le personnel du programme et l'évaluateur peuvent décider de réserver le budget initial pour la conception d'une évaluation bien structurée, l'enquête et l'analyse de base. À mesure que des fonds supplémentaires sont mis à disposition, on peut ajouter au processus global d'évaluation la collecte de données pour l'enquête de suivi et son analyse. Dans les faits, il est plus facile d'obtenir des fonds supplémentaires pour financer l'enquête de suivi que d'obtenir un financement externe pour l'ensemble de l'évaluation.

ANNEXE 3. Tests de vérification et de falsification

Procédures par tirage au sort et par étapes

Les méthodes de répartition aléatoire sont les techniques les plus fiables pour l'estimation du contrefactuel. Elles sont considérées être la référence en la matière pour les évaluations d'impact. Il conviendra d'effectuer certains tests pour évaluer la validité de cette stratégie d'évaluation dans un contexte donné.

- Les caractéristiques de base sont-elles équilibrées ? Comparer les caractéristiques de base du groupe de traitement et du groupe témoin.
- L'assignation est-t-elle non-conforme ? Vérifier si toutes les unités remplissant les conditions requises ont reçu le traitement et si des unités qui ne les remplissent pas l'ont également reçu. Si une assignation est non-conforme, utiliser la méthode randomisée.
- Est-ce que le nombre d'unités du groupe de traitement et du groupe témoin est suffisamment important ? Si ce n'est pas le cas, il faudrait combiner la méthode randomisée avec la méthode de la double différence.

Promotion aléatoire

La promotion aléatoire permet d'obtenir des estimations valables du contrefactuel si la campagne de promotion augmente de façon considérable la participation au programme sans que cela altère directement les résultats d'intérêt.

- Les caractéristiques de base de ceux qui ont bénéficié de la campagne de promotion et ceux qui ne l'ont pas fait sont-elles équilibrées? Comparer les caractéristiques de base des deux groupes.
- Est-ce que la campagne de promotion a eu des effets substantiels sur la participation au programme ? Cela devrait être le cas. Pour le confirmer, il faut comparer le taux de participation du groupe encouragé et celui du groupe non-encouragé.
- La campagne de promotion a-t-elle des effets directs sur les résultats ? Cela ne devrait pas être le cas. Normalement, on ne peut pas vérifier cela directement. Il faudra se fier à la théorie et au bon sens.

Approche par discontinuité

L'approche par discontinuité de la régression requiert que l'indice d'admissibilité soit continu autour du seuil éliminatoire, et que les unités soient comparables autour de ce seuil.

- L'indice est-il continu autour du seuil au moment de l'enquête de base ?
- Le traitement du seuil souffre-t-il de non-conformité ? Vérifier si toutes les unités éligibles et non-éligibles ont reçu le traitement. S'il y a non-conformité, nous devrons combiner l'approche par discontinuité de la régression avec des techniques plus avancées pour corriger cette « discontinuité floue ».

Double différence

La procédure de la double différence repose sur l'hypothèse que les tendances des résultats du groupe de traitement et du groupe de comparaison sont similaires avant l'intervention, et que les facteurs responsables des changements de résultats entre les deux groupes sont constants dans le temps.

- Les résultats auraient-ils évolué en tandem dans les groupes de traitement et le groupe témoin

en absence du programme ? Cela peut être vérifié en utilisant différents tests de falsification, comme (1) Les résultats du groupe de traitement et de comparaison évoluent-ils en tandem avant l'intervention ? Si deux cycles de collecte de données sont disponibles avant la mise en œuvre du programme, vérifier s'il y a des différences de tendance entre les deux groupes. (2) Que fait-on avec les résultats fictifs qui ne devraient pas être influencés par le programme ? Évoluent-ils en tandem avant et après l'intervention dans le groupe de traitement et le groupe témoin ?

- Effectuer l'analyse de la double différence en utilisant plusieurs groupes de comparaison plausibles. Obtenons-nous des estimations similaires concernant l'impact du programme ?
- Effectuer l'analyse de la double différence en utilisant le groupe de traitement et le groupe témoin sélectionnés et un résultat fictif qui ne devrait pas être influencé par le programme. Nous devrions obtenir un impact zéro du programme concernant ce résultat.
- Effectuer l'analyse de la double différence en utilisant la variable du résultat sélectionné pour deux groupes qui ne sont pas touchés par le programme. Nous devrions obtenir un impact zéro du programme.

Appariement

La méthode de l'appariement repose sur l'hypothèse que les unités inscrites et les unités non-inscrites sont similaires pour ce qui est des variables non-observables qui pourraient avoir des effets sur la probabilité de participer au programme ainsi que son résultat.

- La participation au programme est-elle influencée par des variables non-observables ? Si cela ne peut être vérifié directement, il faudra se fier à la théorie et au bon sens.
- Les caractéristiques observées sont-elles équilibrées entre les sous-groupes appariés? Comparer les caractéristiques observables de chaque groupe d'unités de traitement avec son groupe de comparaison apparié.
- Peut-on trouver une unité témoin appariée pour chaque unité de traitement ? Vérifier s'il y a assez de support commun dans la distribution des coefficients de propension. Des aires de support commun réduites indiquent que les personnes inscrites sont très différentes des non-inscrites, et cela met en doute la crédibilité de la méthode d'appariement.

Source : Reproduit avec l'autorisation de Gertler *et al.* (2011, pp. 118–119).

GLOBAL PARTNERSHIP FOR YOUTH EMPLOYMENT

En 2008, avec le support de *World Bank Development Grant Facility,* la Fondation internationale pour la jeunesse (IYF), le Réseau pour l'emploi des jeunes (YEN), *Arab Urban Development Institute,* et *Understanding Children's Work Project* se sont unis afin de créer le *Global Partnership for Youth Employment* (GPYE). Sa mission : construire et diffuser des éléments de preuve faisant état des résultats d'initiatives en emploi jeunesse et de programmes efficaces afin de mieux faire face aux défis auxquels les jeunes sont confrontés dans leur transition vers l'emploi. Le GPYE exploite l'expérience technique et régionale des cinq organisations partenaires quant à la recherche en emploi jeunesse, la programmation, l'évaluation et le dialogue politique. Le travail issu de ce partenariat est axé sur l'Afrique et le Moyen-Orient, régions présentant le besoin d'être mieux informées au sujet d'approches efficaces dans la promotion de l'emploi jeunesse.

Pour plus d'informations, visitez le *www.gpye.org*